

B1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

Case 20918

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
12 December 2002 (12.12.2002)

PCT

(10) International Publication Number
WO 02/099095 A2(51) International Patent Classification⁷: **C12N 9/02**, 9/12, 9/88, 15/53, 15/54, 15/60, 15/61, 1/20, 1/21, C12P 7/02, 17/06

(21) International Application Number: PCT/EP02/06171

(22) International Filing Date: 5 June 2002 (05.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/296,299 6 June 2001 (06.06.2001) US(71) Applicant (for all designated States except US): **ROCHE VITAMINS AG [CH/CH]**; 124 Grenzacherstrasse, CH-4070 Basle (CH).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BERRY, Alan [US/CH]**; 99 Mattweg, CH-4144 Arlesheim (CH). **BRETZEL, Werner [DE/DE]**; 93B Tuellinger Strasse, 79539 Loerrach (DE). **HUEMBELIN, Markus [CH/CH]**; 2 Wollbacherstrasse, CH-4058 Basle (CH). **LOPEZ-ULIBARRI, Rual [MX/CH]**; 3 Wasserweg, CH-4334 Sisseln (CH). **MAYER, Anne Françoise [LU/CH]**; 100 Rennweg, CH-4052 Basle (CH). **YELISEEV, Alexei [RU/US]**; 1076 Carol Lane, # 118, Lafayette, CA 94549 (US).(74) Agent: **HEIROTH, Ulrike**; 124 Grenzacherstrasse, CH-4070 Basle (CH).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report
- with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

**WO 02/099095 A2**

(54) Title: IMPROVED ISOPRENOID PRODUCTION

(57) Abstract: Isolated polynucleotides encoding polypeptides having the activity of enzymes in the mevalonate pathway, e.g. hydroxymethylglutaryl-CoA reductase, isopentenyl diphosphate isomerase, hydroxymethylglutaryl-CoA synthase, mevalonate kinase, phosphomevalonate kinase, or diphosphomevalonate decarboxylase; are provided, useful for recombinantly producing isoprenoid compounds such as carotenoids like phytoene, lycopene, β-carotene, zeaxanthin, canthaxanthin, astaxanthin, adonixanthin, cryptoxanthin, echinenone and adonirubin. Expression vectors, cultured cells, and methods of making isoprenoid compounds are also provided.

Improved Isoprenoid Production

The present invention relates to novel polynucleotides and polypeptide sequences useful in the isoprenoid biosynthetic pathway. More particularly, the present invention provides recombinantly produced cells that exhibit improved production of zeaxanthin. Methods of 5 making and using such cell lines are also provided.

Carotenoids are commercially important C-40 isoprenoid compounds used as nutritional supplements, pharmaceuticals and food colorants for humans and as pigments for animal feed. Currently industrially important carotenoids are produced mainly by chemical synthesis (β -carotene, canthaxanthin and astaxanthin) or extraction from natural sources 10 (lutein from marigold, capsanthin from paprika). Production of carotenoids, however, using microorganisms has been achieved in some cases. For example, β -carotene is produced by fermentation with the fungus *Blakeslea trispora* (US 5,328,845) or by pond culture using the halotolerant alga *Dunaliella salina* [Borowitzka, J. Biotechnol. 70:313-321 15 (1999)]. Lycopene production has also been reported in *B. trispora* (WO 00/77234). Astaxanthin is produced by fermentation using yeast (*Phaffia rhodozyma*, (recently renamed *Xanthophyllomyces dendrorus*)) (US 6,015,684) or in photobioreactors or open ponds using the alga *Haematococcus pluvialis* [Lorenz and Cysewski, Trends Biotechnol. 18:160-167 (1999); Olaizola, J. Appl. Phycol. 12:499-506 (2000)]. Such microbial production systems, however, do not produce carotenoids in amounts sufficient for economic 20 industrial scale production.

In the mid-1960's, scientists at Hoffmann-La Roche isolated several marine bacteria that produced the yellow carotenoid zeaxanthin, which has application in poultry pigmentation and in the prevention of age-related macular degeneration in humans. One bacterium, which showed promising levels of zeaxanthin production, was given the strain designation 25 R-1512, and it was deposited at the American Type Culture Collection (ATCC, Manassas, VA, USA) as strain ATCC 21588 (US 3,891,504). Using the accepted taxonomic standards of that time (classification performed by the Eidgenössische Technische Hochschule (Zürich) and the National Collection of Industrial Bacteria, Torry Research Station (Aberdeen, Scotland)), the zeaxanthin-producing organism was classified as a member of 30 the genus *Flavobacterium*, but no species designation was assigned.

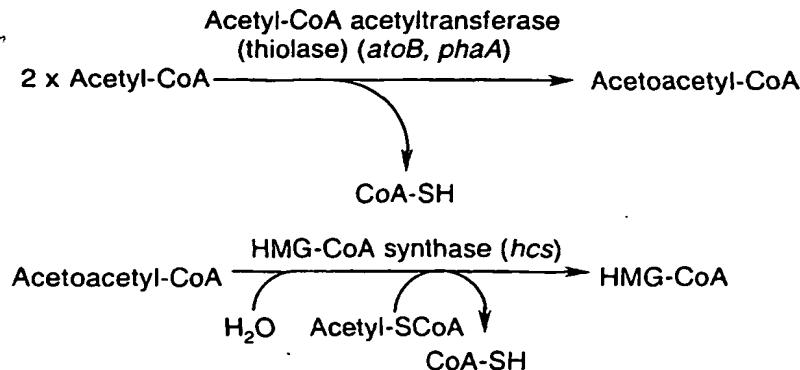
- 2 -

An extensive mutagenesis and screening program was subsequently conducted to isolate mutants of R-1512 with higher zeaxanthin productivities. With respect to the presently described work, two such mutants are significant. These mutants, listed in order of their zeaxanthin productivities, are R1534 and R114. A variety of other mutants have been used 5 over the years for biochemical studies of carotenoid biosynthesis [Goodwin, Biochem. Soc. Symp. 35:233-244 (1972); McDermott et al., Biochem. J. 134:1115-1117 (1973); Britton et al., Arch. Microbiol. 113:33-37 (1977); Mohanty et al., Helvetica Chimica Acta 83:2036-2053 (2000)].

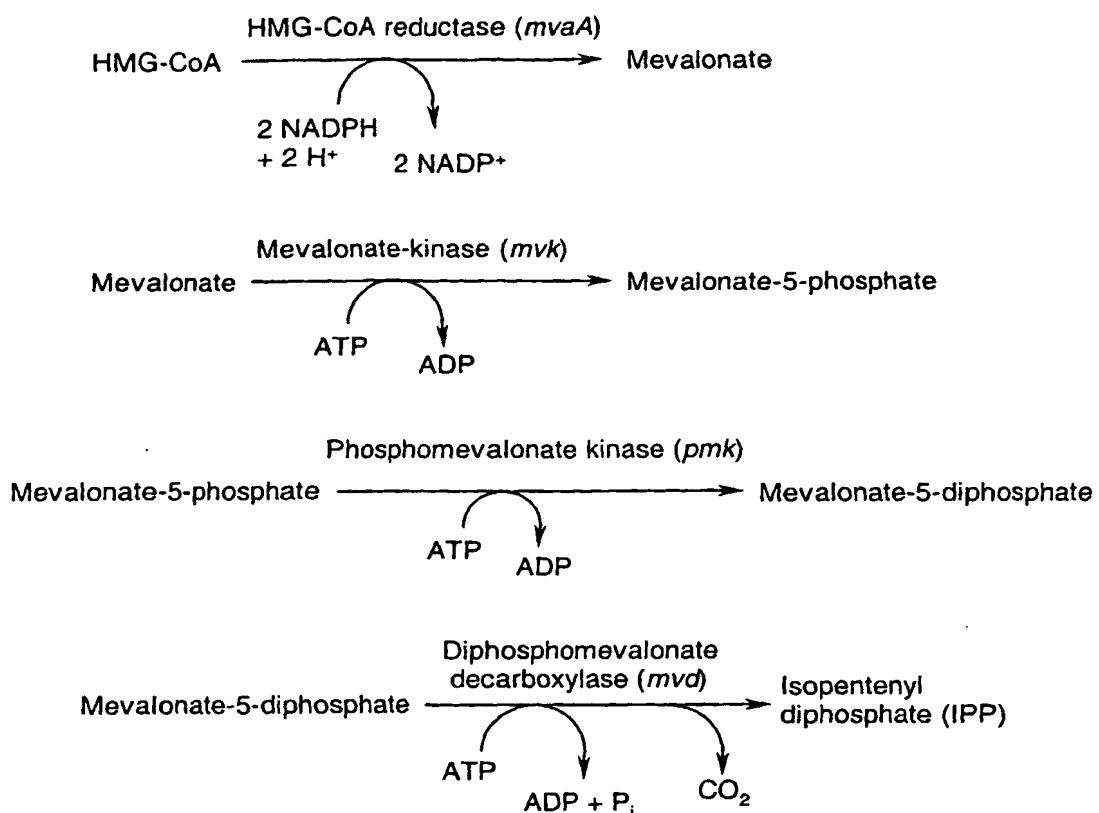
The early attempts to develop a commercially viable fermentation process for the production of zeaxanthin using classically derived mutants of strain R-1512 were not successful. 10 However, with the advent of molecular biology, the possibility arose that higher zeaxanthin-producing strains could be developed. The first step in this direction was taken with the cloning and sequencing of the carotenoid gene cluster from strain R1534 (US 6,087,152), which is hereby incorporated by reference as if recited in full herein). 15 US 6,087,152 discloses that the carotenoid genes were functionally expressed in *Escherichia coli* and *Bacillus subtilis* resulting in zeaxanthin production in these hosts. US 6,087,152 also disclosed that by modifying the carotenoid gene cluster or by adding a gene from an astaxanthin producing bacterium, it was possible to produce carotenoids other than zeaxanthin (EP 872,554). Moreover, EP 872,554 disclosed that carotenoid production was 20 increased in strain R1534 by introducing cloned carotenoid gene clusters on a multi-copy plasmid.

Despite the enormous structural diversity in isoprenoid compounds, all are biosynthesized from a common C-5 precursor, isopentenyl pyrophosphate (IPP). Up until the early 1990's it was generally accepted that IPP was synthesized in all organisms via the mevalonate pathway, even though some experimental results were not consistent with this biogenic 25 scheme [Eisenreich et al., Chemistry and Biology 5:R221-R233 (1998)].

Mevalonate pathway:

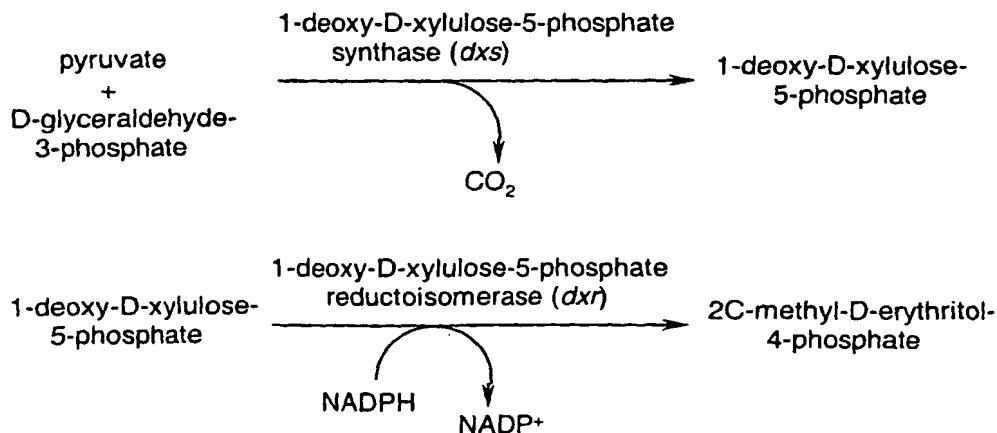


- 3 -

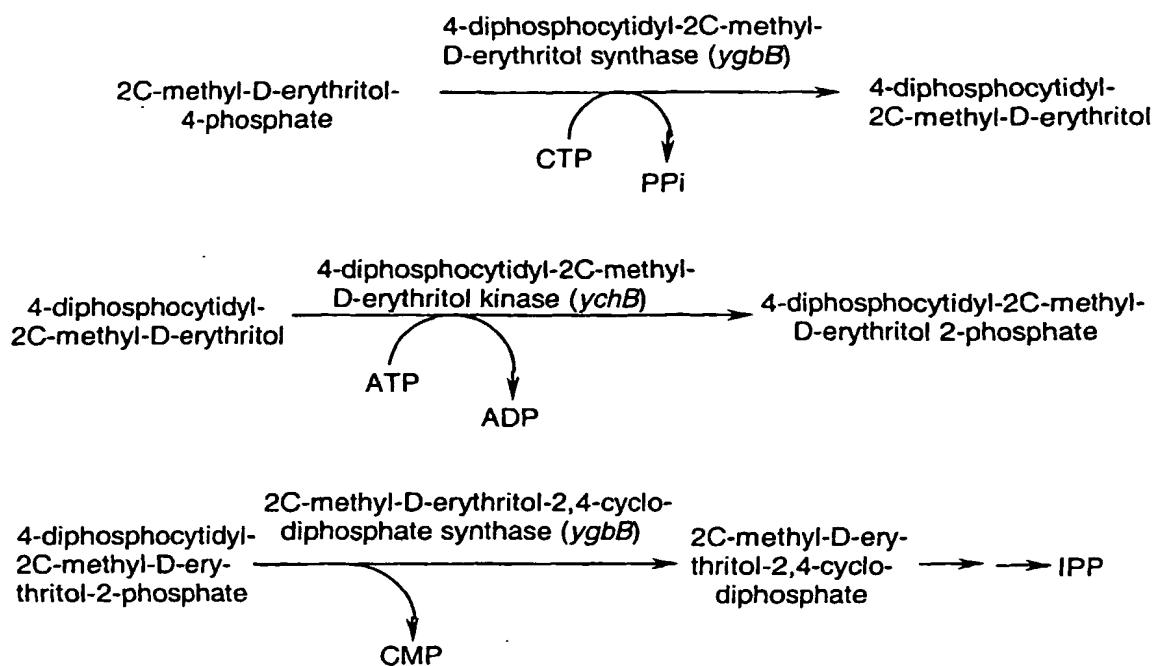


- 5 The discrepancies have since been reconciled by the discovery of an alternate pathway of IPP biosynthesis, the deoxyxylulose (DXP) pathway (Note: The alternate pathway of IPP biosynthesis has been referred to by various names in the scientific literature (DXP pathway, DOXP pathway, MEP pathway, GAP/pyruvate pathway and the non-mevalonate pathway). We use the name DXP pathway here only for the sake of simplicity). The first
 10 five reactions of the DXP pathway have been identified [Herz et al., Proc. Nat. Acad. Sci. 97:2486-2490 (2000)], but the subsequent steps leading to formation of IPP have not yet been elucidated.

DXP pathway:



- 4 -



McDermott et al. (supra) and Britton et al. [J. Chem. Soc. Chem. Comm. p. 27 (1979)]

- 5 showed that crude extracts of zeaxanthin producing mutant strains derived from the original Roche isolates incorporated labeled mevalonate into zeaxanthin. While there was no reason to question this evidence for IPP biosynthesis via the mevalonate pathway, the work was done prior to the discovery of the DXP pathway, and it has been reported that some bacteria (*Streptomyces* species) possess both pathways for IPP synthesis and that expression
- 10 of these pathways is temporally regulated [Seto et al., Tetrahedron Lett. 37:7979-7982 (1996); Dairi et al., Mol. Gen. Genet. 262:957-964 (2000)]. In addition, at present, only a small number of eubacteria have been shown to possess the mevalonate pathway for IPP synthesis. The genes encoding the enzymes of the mevalonate pathway have been cloned and sequenced from some of these bacteria [Wilding et al., J. Bacteriol. 182:4319-4327
- 15 (2000); Takagi et al., J. Bacteriol. 182:4153-4157 (2000)].

Several examples exist where the application of metabolic engineering has succeeded in altering or improving carotenoid production in microorganisms [Lagarde et al., Appl. Env. Microbiol. 66:64-72 (2000); Wang et al., Biotechnol. Bioeng. 62:235-241 (1999); Wang et al., Biotechnol. Prog. 16:922-926 (2000) (and references therein); Sandmann et al., Trends Biotechnol. 17:233-237 (2000); Misawa and Shimada, J. Biotechnol. 59:169-181 (1998); Matthews and Wurtzel, Appl. Microbiol. Biotechnol. 53:396-400 (2000); Albrecht et al., Nature Biotechnol. 18:843-846 (2000); Schmidt-Dannert et al., Nature Biotechnol. 18:750-753 (2000)]. For example, *E. coli*, a non-carotenogenic bacterium, can be engineered to produce carotenoids by introducing the cloned carotenoid (*crt*) genes from the bacteria

- 5 -

Agrobacterium aurantiacum, *Erwinia herbicola* or *Erwinia uredovora* (Misawa and Shimada, supra). Harker and Bramley [FEBS Lett. 448:115-119 (1999)] and Matthews and Wurtzel (supra) disclosed that carotenoid production in such engineered *E. coli* strains could be increased by over-expressing the gene coding for 1-deoxy-D-xylulose 5-phosphate synthase (DXPS), the first enzyme in the DXP pathway (*E. coli* possesses only the DXP pathway for isoprenoid biosynthesis and does not use the mevalonate pathway [Lange et al., Proc. Nat. Acad. Sci. 97:13172-13177 (2000)]. Harker and Bramley (supra) also disclosed an increase in the isoprenoid compound ubiquinone-8, in the cells overproducing DXPS. These results supported the hypothesis that limited availability of IPP, resulting from insufficient *in vivo* activity of DXPS, was limiting the production of carotenoids and other isoprenoid compounds in the engineered strains. Using a similar *E. coli* system, Kim and Keasling [Biotechnol. Bioeng. 72:408-415 (2001)] disclosed that the combined over-expression of the genes encoding DXPS and the second enzyme of the DXP pathway, DXP reductoisomerase (1-deoxy-D-xylulose-5-phosphate reductoisomerase) gave higher carotenoid production than over-expression of just the gene encoding DXPS.

All of these studies were done in *E. coli* engineered to produce carotenoids. Accordingly, one disadvantage to these studies was that the amount of carotenoids produced by these recombinant *E. coli* strains were very low compared to the amounts produced by even non-recombinant microorganisms used for industrial production of carotenoids. Furthermore, improved carotenoid production in bacteria by genetic engineering of the IPP biosynthetic pathway has only been shown in organisms that utilize the DXP pathway for IPP formation. No similar studies have been reported for bacteria that produce IPP via the mevalonate pathway.

Metabolic engineering of the mevalonate pathway to improve production of isoprenoid compounds has been reported in yeast. For example, WO 00/01649 disclosed that production of isoprenoid compounds is increased in *Saccharomyces cerevisiae* when the gene coding for 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA reductase) is over-expressed. However, it has not been shown that this strategy improves isoprenoid production in bacteria, and in particular, it has not been shown that carotenoid production in bacteria can be improved by amplifying expression of mevalonate pathway genes. While it has been shown that some mevalonate pathway genes from eukaryotes [Campos et al., Biochem. J. 353:59-67 (2001)] and from the bacterium *Streptomyces* sp. strain CL190 (Takagi et al., supra) can be expressed in *E. coli*, no increase in isoprenoid production was reported in the strains.

- 6 -

In addition to the reactions that form IPP (via the DXP or mevalonate pathways) and the reactions that convert farnesyl pyrophosphate (FPP) to various other isoprenoids (e.g., carotenoids, quinones) two other reactions are known to be involved in isoprenoid biosynthesis. IPP isomerase interconverts IPP and its isomer, dimethylallyl pyrophosphate (DMAPP). Two forms of IPP isomerase exist, the type 1 enzyme is well known in eukaryotes and some bacteria, and the newly identified type 2 enzyme that is FMN- and NADP(H)-dependent [Kaneda et al., Proc. Nat. Acad. Sci. 98:932-937 (2001)].

Several reports disclose that in *E. coli* engineered to produce carotenoids, amplification of native or heterologous type 1 IPP isomerase (*idi*) genes stimulates carotenoid production 10 [Kajiwara et al., Biochem. J. 324:421-426 (1997); Verdoes and van Ooyen, Acta Bot. Gallica 146:43-53 (1999); Wang et al., supra]. In one report (Wang et al, supra), it was further disclosed that over-expression of the *ispA* gene, encoding FPP synthase (Farnesyl diphosphate synthase) increased carotenoid production in an engineered carotenogenic strain of *E. coli* when combined with over-expression of the *idi* and *crtE* (GGPP synthase/Geranyl-15 geranyl diphosphate synthase) genes. As is the case for the pathway of IPP biosynthesis, however, it has not been shown that over-expression of genes coding for IPP isomerase or FPP synthase improves carotenoid production in a naturally carotenogenic microorganism. Also, the levels of carotenoids produced in the *E. coli* strains described above are very low, and it has not been shown that these strategies work in an industrial microorganism where carotenoid production was already high.

In sum, there is no prior evidence that increased expression of gene(s) coding for enzymes of the mevalonate pathway can improve production of carotenoids in naturally carotenogenic bacteria or in naturally non-carotenogenic bacteria engineered to be carotenogenic.

One embodiment of the present invention is an isolated polypeptide that includes an 25 amino acid sequence selected from the following group: (a) an amino acid sequence shown as residues 1 to 340 of SEQ ID NO:43; (b) an amino acid sequence shown as residues 1 to 349 of SEQ ID NO:45; (c) an amino acid sequence shown as residues 1 to 388 of SEQ ID NO:47; (d) an amino acid sequence shown as residues 1 to 378 of SEQ ID NO:49; (e) an amino acid sequence shown as residues 1 to 305 of SEQ ID NO:51; (f) an amino acid 30 sequence shown as residues 1 to 332 of SEQ ID NO:53; (g) a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOs: 43, 45, 47, 49, 51, and 53, wherein said fragment has at least 30 contiguous amino acid residues; (h) an amino acid sequence of a fragment of a polypeptide selected from the group consisting of SEQ ID NOs: 43, 45, 47, 49, 51, and 53, the fragment having the activity of HMG-CoA reductase; 35 isopentenyl diphosphate isomerase, hydroxymethylglutaryl-CoA synthase (HMG-CoA

- 7 -

synthase), mevalonate kinase, phosphomevalonate kinase, or diphosphomevalonate de-carboxylase; (i) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe comprising at least 30 consecutive nucleotides of SEQ ID NO:42 or a complement of SEQ ID NO:42, wherein the 5 polypeptide has the activity of HMG-CoA reductase, isopentenyl diphosphate isomerase, HMG-CoA synthase, isopentenyl diphosphate isomerase, mevalonate kinase, phosphomevalonate kinase, or diphosphomevalonate decarboxylase; and (j) a conservatively modified variant of SEQ ID NOS:43, 45, 47, 49, 51 or 53.

As noted above, the present invention includes SEQ ID Nos: 43, 45, 47, 49, 51, and 53, 10 which are polypeptide sequences that correspond to the following enzymes of the mevalonate pathway: hydroxymethyl glutaryl CoA (HMG-CoA) reductase, isopentenyl di-phosphate (IPP) isomerase, HMG-CoA synthase, mevalonate kinase, phosphomevalonate kinase, and diphosphomevalonate decarboxylase, respectively. The present invention also includes at least 30 contiguous amino acids of each identified sequence or a sufficient 15 number of contiguous amino acids to define a biologically active molecule.

The present invention also includes fragments of a polypeptide selected from SEQ ID NOS: 43, 45, 47, 49, 51, and 53. The fragment should be at least about 30 amino acids in length but must have the activity of the identified polypeptide, e.g., in the case of SEQ ID NO:43, a fragment thereof that falls within the scope of the present invention has the activity of 20 HMG-CoA reductase. As used herein, a measure of activity of the respective fragments is set forth in Example 1. A fragment having an activity above background in the assays set forth in Example 1 is considered to be biologically active and within the scope of the present invention.

The present invention also includes an amino acid sequence of a polypeptide encoded by a 25 polynucleotide that hybridizes under stringent conditions, as defined above, to a hybridization probe that contains at least 30 contiguous nucleotides of SEQ ID NO:42 (i.e., the mevalonate operon) or a complement of SEQ ID NO:42. The polynucleotide must encode at least one of the enzymes in the mevalonate pathway. For purposes of the present invention, a "hybridization probe" is a polynucleotide sequence containing from about 10- 30 9066 nucleotides of SEQ ID NO:42.

In this embodiment, the isolated polypeptide may have the amino acid sequence of SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51 or SEQ ID NO:53. Alternatively, the isolated polypeptide may contain about 30 contiguous amino acids selected from an area of the respective amino acids sequences that have the least

- 8 -

identity when compared to an enzyme with the same function from different species. Thus, for example, a polypeptide of the present invention may include amino acids 68-97 of SEQ ID NO:43, 1-30 of SEQ ID NO:45, 269-298 of SEQ ID NO:47, 109-138 of SEQ ID NO:49, 198-227 of SEQ ID NO:51 or 81-110 of SEQ ID NO:53.

- 5 Another embodiment of the invention is an isolated polypeptide having an amino acid sequence selected from: (a) an amino acid sequence shown as residues 1 to 287 of SEQ ID NO:159; (b) at least 30 contiguous amino acid residues of SEQ ID NO:159; (c) an amino acid sequence of a fragment of SEQ ID NO:159, the fragment having the activity of farnesyl-diphosphate synthase (FPP synthase); (d) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe containing at least 30 consecutive nucleotides of the *ispA* gene (*i.e.*, nucleotides 295-1158 of SEQ ID NO:157) or a complement thereof, wherein the polypeptide has the activity of FPP synthase; and (e) conservatively modified variants of SEQ ID NO:159.
- 10

Thus, in this embodiment the amino acid may be encoded by the entire open reading frame that encodes FPP synthase, *i.e.*, residues 1-287 of SEQ ID NO:159, at least 30 contiguous residues thereof, or a fragment of SEQ ID NO:159 that has FPP synthase activity as measured by the assay set forth in Example 1. Furthermore, this embodiment of the invention also includes amino acid sequence(s) encoded by polynucleotide(s) that hybridize under stringent conditions, as defined above, to a hybridization probe that includes at least 20 30 consecutive nucleotides of the *ispA* gene (*i.e.*, nucleotides 295-1158 of SEQ ID NO:157) or a complement thereof, wherein the polypeptide has FPP synthase activity as defined above.

In a preferred embodiment, the polypeptide has the amino acid sequence of SEQ ID NO:159.

- 25 Another embodiment of the invention is an isolated polypeptide having an amino acid sequence selected from the following group: (a) an amino acid sequence shown as residues 1 to 142 of SEQ ID NO:160; (b) at least 30 contiguous amino acid residues of SEQ ID NO:160; (c) an amino acid sequence of a fragment of SEQ ID NO: 160, the fragment having the activity of 1-deoxyxylulose-5-phosphate synthase (DXPS); (d) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe containing at least 30 consecutive nucleotides spanning positions 1185-1610 of SEQ ID NO:157 or a complement thereof, wherein the polypeptide has the activity of DXPS; and (e) conservatively modified variants of SEQ ID NO:160.
- 30

- 9 -

Thus, in this embodiment the amino acid may be encoded by the entire open reading frame that encodes DXPS, *i.e.*, residues 1-142 of SEQ ID NO:160, at least 30 contiguous residues thereof, or a fragment of SEQ ID NO:160 that has DXPS activity as measured by as measured by the assay set forth in Example 1. Furthermore, this embodiment of the 5 invention also includes amino acid sequence(s) encoded by polynucleotide(s) that hybridize under stringent conditions, as defined above, to a hybridization probe that includes at least 30 consecutive nucleotides of the DXPS gene (*i.e.*, nucleotides 1185-1610 of SEQ ID NO:157) or a complement thereof, wherein the polypeptide has DXPS activity as defined above.

- 10 In a preferred embodiment, the polypeptide has the amino acid sequence of SEQ ID NO:160.

Another embodiment of the invention is an isolated polypeptide having an amino acid sequence selected from: (a) an amino acid sequence shown as residues 1 to 390 of SEQ ID NO:178; (b) at least 30 contiguous amino acid residues of SEQ ID NO:178; (c) an amino 15 acid sequence of a fragment of SEQ ID NO:178, the fragment having the activity of acetyl-CoA acetyltransferase; (d) an amino acid sequence of a polypeptide encoded by a poly-nucleotide that hybridizes under stringent conditions to a hybridization probe containing at least 30 consecutive nucleotides of the *phaA* gene (*i.e.*, nucleotides 1-1179 of SEQ ID NO:177) or a complement thereof, wherein the polypeptide has the activity of acetyl-CoA. 20 acetyltransferase, and (e) conservatively modified variants of SEQ ID NO:178.

Thus, in this embodiment the amino acid may be encoded by the entire open reading frame that encodes acetyl-CoA acetyltransferase, *i.e.*, residues 1-143 of SEQ ID NO:178, at least 30 contiguous residues thereof, or a fragment of SEQ ID NO:178 that has acetyl-CoA acetyltransferase activity as measured by the assay set forth in Example 1. Furthermore, 25 this embodiment of the invention also includes amino acid sequence(s) encoded by polynucleotide(s) that hybridize under stringent conditions, as defined above, to a hybridization probe that includes at least 30 consecutive nucleotides of the *phaA* gene (*i.e.*, nucleotides 1-1170 of SEQ ID NO:177), or a complement thereof, wherein the polypeptide has the acetyl-CoA acetyltransferase activity as defined above.

- 30 In a preferred embodiment, the polypeptide has the amino acid sequence of SEQ ID NO:178.

Another embodiment of the invention is an isolated polypeptide having an amino acid sequence selected from: (a) an amino acid sequence shown as residues 1 to 240 of SEQ ID

- 10 -

- NO:179; (b) at least 30 contiguous amino acid residues of SEQ ID NO:179; (c) an amino acid sequence of a fragment of a polypeptide of SEQ ID NO:179, the fragment having the activity of acetoacetyl-CoA reductase; (d) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe containing at least 30 consecutive nucleotides of the *phaB* gene (*i.e.*, nucleotides 1258-1980 of SEQ ID NO:177) or a complement thereof, wherein the polypeptide has the activity of acetoacetyl-CoA reductase; and (e) conservatively modified variants of SEQ ID NO:179.

Thus, in this embodiment the amino acid may be encoded by the entire open reading frame that encodes acetoacetyl-CoA reductase, *i.e.*, residues 1-240 of SEQ ID NO:179, at least 30 contiguous residues thereof, or a fragment of SEQ ID NO:179 that has acetoacetyl-CoA reductase activity as measured by the assay set forth in Example 1. Furthermore, this embodiment of the invention also includes amino acid sequence(s) encoded by polynucleotide(s) that hybridize under stringent conditions, as defined above, to a hybridization probe that includes at least 30 consecutive nucleotides of the *phaB* gene (*i.e.*, nucleotides 1258-1980 of SEQ ID NO:177) or a complement thereof, wherein the polypeptide has acetoacetyl-CoA reductase activity as defined above.

In a preferred embodiment, the polypeptide has the amino acid sequence of SEQ ID NO:179.

- The terms "polypeptide," "polypeptide sequence," "amino acid," and "amino acid sequence" are used interchangeably herein, and mean an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, as well as naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of any of the polypeptides defined herein which are at least about 30 amino acids in length and which retain some biological activity or immunological activity of the polypeptide in question. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.
- With respect to polypeptides, the term "isolated" means a protein or a polypeptide that has been separated from components that accompany it in its natural state. A monomeric protein is isolated when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. An isolated protein will typically comprise about 60 to 90% W/W of a protein sample, more usually about 95%, and preferably will be over about 99% pure. Protein

- 11 -

purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes, using HPLC or other means well known in the art may provide higher resolution for purification.

- 5 As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic polypeptide, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.
- 10 Another embodiment of the invention is an isolated polynucleotide sequence having the nucleotide sequence of the mevalonate operon (SEQ ID NO:42), variants of SEQ ID NO:42 containing one or more substitutions according to the *Paracoccus* sp. strain 1534 codon usage table (see Table 14) or fragments of SEQ ID NO:42. The variants and fragments of SEQ ID NO:42 must encode a polypeptide having an activity selected from:
- 15 HMG-CoA reductase, isopentenyl diphosphate isomerase activity, hydroxymethylglutaryl-CoA synthase (HMG-CoA synthase), mevalonate kinase, phosphomevalonate kinase, and diphosphomevalonate decarboxylase. This embodiment also includes polynucleotide sequences that hybridize under stringent conditions, as defined above, to a hybridization probe, the nucleotide sequence of which consists of from about 10 to about 9066 nucleotides of SEQ ID NO:42, preferably at least 30 contiguous nucleotides of SEQ ID NO:42, or a complement of such sequences, which polynucleotide encodes a polypeptide having an activity selected from: HMG-CoA reductase, isopentenyl diphosphate isomerase, HMG-CoA synthase, mevalonate kinase, phosphomevalonate kinase, and diphosphomevalonate decarboxylase.
- 20
- 25 This embodiment also includes isolated polynucleotide sequences spanning the following residues of SEQ ID NO:42: 2622 to 3644, 3641 to 4690, 4687 to 5853, 5834 to 6970, 6970 to 7887, 7880 to 8878. Fragments of these sequences are also within the scope of the invention, so long as they encode a polypeptide having HMG-CoA reductase activity, isopentenyl diphosphate isomerase activity, HMG-CoA synthase activity, mevalonate kinase activity, phosphomevalonate kinase activity, and diphosphomevalonate de-carboxylase activity, respectively.
- 30

This embodiment also includes polynucleotide sequences that hybridize under stringent conditions, as defined above, to a hybridization probe selected from a nucleotide sequence which consists of at least 30 contiguous nucleotides of the following residues of SEQ ID

- 12 -

NO:42: 2622 to 3644, 3641 to 4690, 4687 to 5853, 5834 to 6970, 6970 to 7887, 7880 to 8878 or a complement thereof, wherein the polynucleotide encodes a polypeptide having HMG-CoA reductase activity, isopentenyl diphosphate isomerase activity, HMG-CoA synthase activity, mevalonate kinase activity, phosphomevalonate kinase activity, or diphospho-
5 mevalonate decarboxylase activity, respectively.

Preferably, the isolated polynucleotide consists of nucleotides 2622 to 3644, 3641 to 4690, 4687 to 5853, 5834 to 6970, 6970 to 7887 or 7880 to 8878 of SEQ ID NO:42.

Another embodiment of the invention is an isolated polynucleotide sequence having the nucleotide sequence of SEQ ID NO:157, variants of SEQ ID NO:157 containing one or
10 more substitutions according to the *Paracoccus* sp. strain 1534 codon usage table (see Table 14) or fragments of SEQ ID NO:157 that encode a polypeptide having FPP synthase activity, 1-deoxy-D-xylulose 5-phosphate synthase activity or the activity of XseB. This embodiment also includes polynucleotide sequences that hybridize under stringent conditions, as defined above, to a hybridization probe the nucleotide sequence of which consists
15 of at least 30 contiguous nucleotides of SEQ ID NO:157, or the complement of SEQ ID NO:157, wherein the polynucleotide encodes a polypeptide having FPP synthase activity, 1-deoxy-D-xylulose 5-phosphate synthase activity or the activity of XseB.

Preferably, the isolated polynucleotide consists of nucleotides 59-292, 295-1158 or 1185-1610 of SEQ ID NO:157.

- 20 An isolated polynucleotide sequence is also provided that has a nucleotide sequence selected from the following group: nucleotides spanning positions 59-292 of SEQ ID NO:157, variants of the nucleotide sequence spanning positions of SEQ ID NO:157 containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of the nucleotide sequence spanning positions 59-292 of
25 SEQ ID NO:157 that encode a polypeptide having a function of XseB, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning positions 59-292 of SEQ ID NO:157, or the complement of such a sequence, wherein the polynucleotide encodes a polypeptide having a function of XseB.
- 30 Preferably, the isolated polynucleotide consists of nucleotides 59 to 292 of SEQ ID NO:157.

An isolated polynucleotide sequence is also provided that has a nucleotide sequence selected from the following group: nucleotides spanning positions 295-1158 of SEQ ID

- 13 -

NO:157, variants of the nucleotide sequence spanning positions 295-1158 of SEQ ID NO:157 containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of the nucleotide sequence spanning positions 295-1158 of SEQ ID NO:157 that encode a FPP synthase activity, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning positions 295-1158 of SEQ ID NO:157, or the complement of such a sequence, wherein the polynucleotide encodes a polypeptide having FPP synthase activity.

Preferably, the isolated nucleotide sequence consists of nucleotides 295-1158 of SEQ ID NO:157.

Another embodiment of the invention is an isolated polynucleotide sequence having the nucleotide sequence spanning positions 1185-1610 of SEQ ID NO:157, variants of the nucleotide sequence spanning positions 1185-1610 of SEQ ID NO:157 containing one or more substitutions according to the *Paracoccus* sp. strain 1534 codon usage table (see Table 14) or fragments of the nucleotide sequence spanning positions 1185-1610 of SEQ ID NO:157 that encode a polypeptide having 1-deoxyxylulose-5-phosphate synthase activity. This embodiment also includes polynucleotide sequences that hybridize under stringent conditions, as defined above, to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning positions 1185-1610 of SEQ ID NO:157, or a complement thereof, wherein the polynucleotide encodes a polypeptide having 1-deoxyxylulose-5-phosphate synthase activity.

Preferably, the isolated polynucleotide consists of nucleotides 1185 to 1610 of SEQ ID NO:157.

Another embodiment of the invention is an isolated polynucleotide sequence having the nucleotide sequence of SEQ ID NO:177, variants of SEQ ID NO:177 containing one or more substitutions according to the *Paracoccus* sp. strain 1534 codon usage table (see Table 14) or fragments of SEQ ID NO:177 that encode a polypeptide having an activity selected from acetyl-CoA acetyltransferase and acetoacetyl-CoA reductase. This embodiment also includes polynucleotide sequences that hybridize under stringent conditions, as defined above, to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides of SEQ ID NO:177, or a complement thereof, which polynucleotide encodes a polypeptide having an activity selected from the group consisting of acetyl-CoA acetyltransferase and acetoacetyl-CoA reductase.

- 14 -

- In this embodiment the isolated polynucleotide sequence may include nucleotides 1 to 1170 of SEQ ID NO:177, variants of SEQ ID NO:177 containing one or more substitutions according to the *Paracoccus* sp. strain 1534 codon usage table (see Table 14) or fragments of SEQ ID NO:177 that encode a polypeptide having acetyl-CoA acetyltransferase activity.
- 5 This embodiment also includes polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides of nucleotides 1 to 1170 of SEQ ID NO:177, or a complement thereof, wherein the polynucleotide encodes a polypeptide having acetyl-CoA acetyl-transferase activity.
- 10 Preferably, the isolated polynucleotide sequence consists of nucleotides 1-1170 of SEQ ID NO:177.
- In this embodiment, the isolated polynucleotide sequence may alternatively be nucleotides 1258-1980 of SEQ ID NO:177, variants of SEQ ID NO:177 containing one or more substitutions according to the *Paracoccus* sp. strain 1534 codon usage table (see Table 14) or
- 15 fragments of SEQ ID NO:177 that encode a polypeptide having acetoacetyl-CoA reductase activity. This embodiment also includes polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides of nucleotides 1258-1980 of SEQ ID NO:177, or a complement thereof, wherein the polynucleotide encodes a polypeptide having acetoacetyl-
- 20 CoA reductase activity.
- Preferably, the isolated polynucleotide consists of nucleotides 1258-1980 of SEQ ID NO:177.
- In another embodiment of the invention, the isolated polynucleotide sequence has a nucleotide sequence selected from SEQ ID NO:42, SEQ ID NO:157, SEQ ID NO:177, and
- 25 combinations thereof. As used herein, the phrase "and combinations thereof" when used in reference to nucleotide sequences means that any combination of the recited sequences may be combined to form the isolated polynucleotide sequence. Moreover, in the present invention, multiple copies of the same sequence, i.e., concatamers may be used. Likewise, and as set forth in more detail below, multiple copies of plasmids containing the same
- 30 polynucleotide sequence may be transferred into suitable host cells.

As used herein, an "isolated" polynucleotide (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components which naturally accompany a native sequence or polypeptide, e.g., ribosomes, polymerases, many other

- 15 -

genome sequences and proteins. The term embraces a polynucleotide that has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs or analogs biologically synthesized by heterologous systems.

- 5 The phrase "nucleic acid sequence" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It includes chromosomal DNA, self-replicating plasmids, infectious polymers of DNA or RNA and DNA or RNA that performs a primarily structural role.

An "expression control sequence" is defined as an array of nucleic acid control sequences 10 that direct transcription of an operably linked nucleic acid. An example of such an expression control sequence is a "promoter." Promoters include necessary nucleic acid sequences near the start site of transcription. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is 15 active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the 20 nucleic acid corresponding to the second sequence.

A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the 25 promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants.

In the case of both expression of transgenes and inhibition of endogenous genes (e.g., by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only "substantially identical" to a sequence of 30 the gene from which it was derived.

In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These vari-

- 16 -

ants are specifically within the scope of the present invention. In addition, the present invention specifically includes those sequences that are substantially identical (determined as described below) to each other and that encode polypeptides that are either mutants of wild type polypeptides or retain the function of the polypeptide (e.g., resulting from conservative substitutions of amino acids in the polypeptide). In addition, variants can be those that encode dominant negative mutants as described below.

- Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or 10 percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual 15 inspection. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences 20 differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a 25 score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller, Computer Applic. Biol. Sci. 4:11-17 (1988), e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, Calif., USA).
- 30 The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to sequences or subsequences that have at least 60%, preferably 80%, most preferably 90-95%, nucleotide or amino acid residue identity when aligned for maximum correspondence over a comparison window as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. This

definition also refers to a sequence of which the complement of that sequence hybridizes to the test sequence.

- For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and
- 5 reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.
- 10 A "comparison window," as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for com-
- 15 parison are well known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algo-
- 20 rithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection.
- One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show
- 25 relationship and percent sequence identity. It also plots a tree or dendrogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng and Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins and Sharp, CABIOS 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000
- 30 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of pro-
- 35 gressive, pairwise alignments. The program is run by designating specific sequences and

- 18 -

their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and
5 weighted end gaps.

Another example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm [Altschul et al., J. Mol. Biol. 215:403-410 (1990)]. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This
10 algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find
15 longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of
20 either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.
25 The BLAST algorithm also performs a statistical analysis of the similarity between two sequences [see, e.g., Karlin and Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)]. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is
30 considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers
35 to those nucleic acids which encode identical or essentially identical amino acid sequences,

- 19 -

or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acid codons encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an
5 alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein that encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid
10 (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid that encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, or substitutions to a peptide, polypeptide, or protein sequence which alters a single amino acid or a small percentage of amino acids (*i.e.* less than 20%, such as 15%, 10%, 5%, 4%, 3%, 2% or 1%) in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.
15

20 The following six groups each contain amino acids that are conservative substitutions for one another:

Alanine (A), Serine (S), Threonine (T);
Aspartic acid (D), Glutamic acid (E);
Asparagine (N), Glutamine (Q);
25 Arginine (R), Lysine (K);
Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
Phenylalanine (F), Tyrosine (Y), Tryptophan (W). (see, *e.g.*, Creighton, Proteins (1984)).

An indication that two nucleic acid sequences or polypeptides are substantially identical is
30 that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two

- 20 -

molecules or their complements hybridize to each other under stringent conditions, as described below.

The phrase "specifically hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization

- 5 conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target sequence, typically in a complex mixture of nucleic acid sequences, but to no other sequences. Stringent conditions are sequence-dependent and will be

- 10 different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen,

Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes, "Overview of Principles of Hybridization and the Strategy of Nucleic Acid Assays" (1993).

Generally, highly stringent conditions are selected to be about 5-10°C lower than the

- 15 thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

Low stringency conditions are generally selected to be about 15-30°C below the T_m . The T_m is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are

- 20 occupied at equilibrium). Stringent conditions will be those in which the salt

concentration is less than about 1.0M sodium ion, typically about 0.01 to 1.0M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition

- 25 of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides that they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon

- 30 degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions.

In the present invention, genomic DNA or cDNA containing nucleic acids of the invention can be identified in standard Southern blots under stringent conditions using the nucleic

- 21 -

acid sequences disclosed here. For the purposes of this disclosure, suitable stringent conditions for such hybridizations are those which include hybridization in a buffer of 40% formamide, 1M NaCl, 1% sodium dodecyl sulfate (SDS) at 37°C, and at least one wash in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C to about 60°C, for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

A further indication that two polynucleotides are substantially identical is if the reference sequence, amplified by a pair of oligonucleotide primers, can then be used as a probe under stringent hybridization conditions to isolate the test sequence from a cDNA or genomic library, or to identify the test sequence in, e.g., a northern or Southern blot.

The present invention also includes expression vectors as defined above. The expression vectors include one or more copies of each of the polynucleotide sequences set forth above. The expression vectors of the present invention may contain any of the polynucleotide sequences defined herein, such as for example SEQ ID NO:42, or the following residues of SEQ ID NO:42: 2622 to 3644, 3641 to 4690, 4687 to 5853, 5834 to 6970, 6970 to 7887, 7880 to 8878, as well as residues 59-292, 295-1158 or 1185-1610 of SEQ ID NO:157 and residues 1-1170 or 1258-1980 of SEQ ID NO:177. The expression vectors may contain combinations of the polynucleotide sequences identified herein, such as for example, SEQ ID NO:42, SEQ ID NO:157, and SEQ ID NO:177.

The polynucleotide sequences in the expression vectors may optionally be operably linked to an expression control sequence as defined above and exemplified in the Examples.

The present invention also includes for example, the following expression vectors: pBBR-K-mev-op16-1, pBBR-K-mev-op16-2, pDS-mvaA, pDS-idi, pDS-hcs, pDS-mvk, pDS-pmk, pDS-mvd, pDS-His-mvaA, pDS-His-idi, pDS-His-hcs, pDS-His-mvk, pDS-His-pmk, pDS-His-mvd, pBBR-K-Zea4, pBBR-K-Zea4-up, pBBR-K-Zea4-down, pBBR-K-P crtE-crtE-3, pBBR-tK-P crtE-mvaA, pBBR-tK-P crtE-idi, pBBR-tK-P crtE-hcs, pBBR-tK-P crtE-mvk, pBBR-tK-P crtE-pmk, pBBR-tK-P crtE-mvd, pBBR-K-P crtE-mvaA-crtE-3, pDS-His-phA, pBBR-K-P crtE-crtW, pBBR-K-P crtE-crtWZ, pBBR-K-P crtE-crtZW, and combinations thereof. These expression vectors are defined in more detail in the examples below. Moreover, the present invention also includes any expression vector that contains one of the sequences defined herein, which expression vector is used to express an isoprenoid compound, such as a carotenoid, preferably zeaxanthin, in a suitable host cell.

- 22 -

As used herein, the phrase "expression vector" is a replicatable vehicle that carries, and is capable of mediating the expression of, a DNA sequence encoding the polynucleotide sequences set forth herein.

- In the present context, the term "replicatable" means that the vector is able to replicate in a given type of host cell into which it has been introduced. Immediately upstream of the polynucleotide sequence(s) of interest, there may be provided a sequence coding for a signal peptide, the presence of which ensures secretion of the encoded polypeptide expressed by host cells harboring the vector. The signal sequence may be the one naturally associated with the selected polynucleotide sequence or of another origin.
- 5 The vector may be any vector that may conveniently be subjected to recombinant DNA procedures, and the choice of vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, i.e. a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication; examples of such a vector are a plasmid, phage, cosmid or mini-10 chromosome. Alternatively, the vector may be one which, when introduced in a host cell, is integrated in the host cell genome and is replicated together with the chromosome(s) into which it has been integrated. Examples of suitable vectors are shown in the examples. The expression vector of the invention may carry any of the DNA sequences of the invention as defined below and be used for the expression of any of the polypeptides of the invention defined below.
- 15 20

The present invention also includes cultured cells containing one or more of the polynucleotide sequences and/or one or more of the expression vectors disclosed herein. As used herein, a "cultured cell" includes any cell capable of growing under defined conditions and expressing one or more of polypeptides encoded by a polynucleotide of the present invention. Preferably, the cultured cell is a yeast, fungus, bacterium, or alga. More preferably, the cultured cell is a *Paracoccus*, *Flavobacterium*, *Agrobacterium*, *Alcaligenes*, *Erwinia*, *E. coli* or *B. subtilis*. Even more preferably, the cell is a *Paracoccus*, such as for example, R-1506, R-1512, R1534 or R114. The present invention also includes the progeny of any of the cells identified herein that express a polypeptide disclosed herein. In the present invention, a cell is a progeny of another cell if its AFLP DNA fingerprint is indistinguishable using the conditions set forth in Example 2 from the fingerprint of the putative parental cell.

25 30

Thus, the cultured cells according to the present invention may contain, for example, SEQ ID NO:42, or the following residues of SEQ ID NO:42: 2622 to 3644, 3641 to 4690, 4687

- 23 -

to 5853, 5834 to 6970, 6970 to 7887, 7880 to 8878, as well as residues 59-292, 295-1158 or 1185-1610 of SEQ ID NO:157 and residues 1-1170 or 1258-1980 of SEQ ID NO:177. These sequences may be transferred to the cells alone or as part of an expression vector. These sequences also may optionally be operatively linked to expression control sequence(s). The cultured cells may also contain combinations of the polynucleotide sequences identified herein, such as for example, SEQ ID NO:42, SEQ ID NO:157, and SEQ ID NO:177.

The cultured cells according to the present invention may further contain polynucleotides that encode one or more enzymes in the carotenoid biosynthetic pathway. For example, 10 the cultured cells according to the present invention may contain one or more copies of SEQ ID NOS:180, 182, and 184 alone or in combination with any of the polynucleotide sequences identified herein. Thus, the polynucleotide sequences disclosed herein may be transferred into a cultured cell alone or in combination with another polynucleotide sequence that would provide enhanced production of the target isoprenoid compound, 15 such as, for example, carotenoids like zeaxanthin or astaxanthin. In this regard, the present invention includes the use of any polynucleotide encoding, for example, a polypeptide involved in carotenoid biosynthesis, such as GGPP synthase, β -carotene- β 4-oxygenase (ketolase), and/or β -carotene hydroxylase. In addition, combinations of polynucleotides encoding polypeptides involved in carotenoid biosynthesis may be used in combination 20 with one or more of the polynucleotides identified herein on the same or different expression vectors. Such constructs may be transferred to a cultured cell according to the present invention to provide a cell that expresses an isoprenoid of interest.

For example, a cultured cell according to the present invention may contain one or more of the following expression vectors: pBBR-K-mev-op16-1, pBBR-K-mev-op16-2, pDS- 25 mvaA, pDS-idi, pDS-hcs, pDS-mvk, pDS-pmk, pDS-mvd, pDS-His-mvaA, pDS-His-idi, pDS-His-hcs, pDS-His-mvk, pDS-His-pmk, pDS-His-mvd, pBBR-K-Zea4, pBBR-K-Zea4-up, pBBR-K-Zea4-down, pBBR-K-PcrtE-crtE-3, pBBR-tK-PcrtE-mvaA, pBBR-tK-PcrtE-idi, pBBR-tK-PcrtE-hcs, pBBR-tK-PcrtE-mvk, pBBR-tK-PcrtE-pmk, pBBR-tK-PcrtE-mvd, pBBR-K-PcrtE-mvaA-crtE-3, pDS-His-phA, pBBR-K-PcrtE-crtW, pBBR-K-PcrtE-crtWZ, 30 pBBR-K-PcrtE-crtZW, and combinations thereof.

Another embodiment of the invention is a method of producing a carotenoid. In this method, a cultured cell as defined above is cultured under conditions permitting expression of a polypeptide encoded by the polynucleotide sequence as defined above. Culture conditions that permit expression of a polypeptide are provided in the Examples

- 24 -

below, but may be modified, if required, to suit the particular intended use. The carotenoid is then isolated from the cell or, if secreted, from the medium of the cell.

In the present invention, a "carotenoid" includes the following compounds: phytoene, lycopene, β -carotene, zeaxanthin, canthaxanthin, astaxanthin, adonixanthin, cryptoxanthin, echinenone, adonirubin, and combinations thereof. Preferably, the carotenoid is zeaxanthin.

Another embodiment of the invention is a method of making a carotenoid-producing cell. This method includes (a) introducing into a cell a polynucleotide sequence encoding an enzyme in the mevalonate pathway, which enzyme is expressed in the cell; and (b) selecting a cell containing the polynucleotide sequence of step (a) that produces a carotenoid at a level that is about 1.1-1,000 times the level of the carotenoid produced by the cell before introduction of the polynucleotide sequence.

As used herein, the phrase "an enzyme in the mevalonate pathway" means the enzymes involved in the mevalonate pathway for IPP biosynthesis and encoded by the *atoB* or *phaA*, *hcs*, *mvaA*, *mvk*, *pmk*, and *mvd* genes. For purposes of the present invention, an enzyme is "expressed in the cell" if it is detected using any one of the activity assays set forth in Example 1. Assays for detecting the production of a carotenoid are well known in the art. Examples 1, 11, and 12 provide typical assay procedures for identifying the presence of zeaxanthin, lycopene, and astaxanthin, respectively. In a similar manner, assays for the other carotenoids may be used to detect the presence in the cell or medium of e.g. phytoene, canthaxanthin, adonixanthin, cryptoxanthin, echinenone, and adonirubin.

Thus, this method may be used to make the following exemplary carotenoids: phytoene, lycopene, β -carotene, zeaxanthin, canthaxanthin, astaxanthin, adonixanthin, cryptoxanthin, echinenone, adonirubin, and combinations thereof. In this method, zeaxanthin is the preferred carotenoid.

This method includes producing cells capable of producing a carotenoid at a level that is about 1.1-1,000 times, preferably about 1.5-500 times, such as about 100 times or at least 10 times, the level of the carotenoid produced by the cell before introduction of the polynucleotide sequence.

In this method, the cell produces from about 1 mg/L to about 10 g/L of a carotenoid. It is preferred that the cell produces from about 100 mg/L to about 9 g/L, such as, for example, from about 500 mg/L to about 8 g/L, or from about 1 g/L to about 5 g/L, of a carotenoid.

- 25 -

In this method, the cell may be selected from a yeast, fungus, bacterium, and alga. Preferably, the cell is a bacterium selected from *Paracoccus*, *Flavobacterium*, *Agrobacterium*, *Alcaligenes*, *Erwinia*, *E. coli*, and *B. subtilis*. More preferably, the bacterium is a *Paracoccus*.

- In this method, the cell may be a mutant cell. As used herein, a "mutant cell" is any cell 5 that contains a non-native polynucleotide sequence or a polynucleotide sequence that has been altered from its native form (e.g., by rearrangement or deletion or substitution of from 1-100, preferably 20-50, more preferably less than 10 nucleotides). Such a non-native sequence may be obtained by random mutagenesis, chemical mutagenesis, UV-irradiation, and the like. Preferably, the mutation results in the increased expression of one 10 or more genes in the mevalonate pathway that results in an increase in the production of a carotenoid, such as zeaxanthin. Methods for generating, screening for, and identifying such mutant cells are well known in the art and are exemplified in the Examples below.

Examples of such mutants are R114 or R1534. Preferably, the mutant cell is R114.

NP 2 mev- op80N

- In this method, the polynucleotide sequence is SEQ ID NO:42, or the following residues of 15 SEQ ID NO:42: 2622 to 3644, 3641 to 4690, 4687 to 5853, 5834 to 6970, 6970 to 7887, 7880 to 8878, as well as residues 59-292, 295-1158 or 1185-1610 of SEQ ID NO:157 and residues 1-1170 or 1258-1980 of SEQ ID NO:177. These sequences may be used in this method alone or as part of an expression vector. These sequences also may optionally be operatively linked to expression control sequence(s). In this method, combinations of the 20 polynucleotide sequences identified herein may be used, such as for example, SEQ ID NO:42, SEQ ID NO:157, and SEQ ID NO:177.

- Examples of expression vector that may be selected for use in this method include pBBR-K-mev-op16-1, pBBR-K-mev-op16-2, pDS-mvaA, pDS-idi, pDS-hcs, pDS-mvk, pDS-pmk, pDS-mvd, pDS-His-mvaA, pDS-His-idi, pDS-His-hcs, pDS-His-mvk, pDS-His-pmk, pDS-His-mvd, pBBR-K-Zea4, pBBR-K-Zea4-up, pBBR-K-Zea4-down, pBBR-K-PcrtE-crtE-3, pBBR-tK-PcrtE-mvaA, pBBR-tK-PcrtE-idi, pBBR-tK-PcrtE-hcs, pBBR-tK-PcrtE-mvk, pBBR-tK-PcrtE-pmk, pBBR-tK-PcrtE-mvd, pBBR-K-PcrtE-mvaA-crtE-3, pDS-His-phA, pBBR-K-PcrtE-crtW, pBBR-K-PcrtE-crtWZ, pBBR-K-PcrtE-crtZW, and combinations thereof.
- 30 In this method, the polynucleotide sequence is introduced into the cell using any conventional means. Examples of suitable methods for introducing a polynucleotide sequence into a cell include transformation, transduction, transfection, lipofection, electroporation [see e.g., Shigekawa and Dower, Biotechniques 6:742-751 (1988)], conjugation [see e.g., Koehler and Thorne, Journal of Bacteriology 169:5771-5278 (1987)], and biolistics.

The use of conjugation to transfer a polynucleotide sequence, such as in the form of an expression vector, into recipient bacteria is generally effective, and is a well-known procedure. (e.g. US 5,985,623). Depending on the strain of bacteria, it may be more common to use transformation of competent cells with purified DNA.

- 5 Known electroporation techniques (both *in vitro* and *in vivo*) function by applying a brief high voltage pulse to electrodes positioned around the treatment region. (e.g. US 6,208,893). The electric field generated between the electrodes causes the cell membranes to temporarily become porous, whereupon molecules of the implant agent enter the cells. In known electroporation applications, this electric field comprises a single
- 10 square wave pulse on the order of 1000 V/cm of about 100 μ s duration. Such a pulse may be generated, for example, in known applications of the Electro Square Porator T820, made by the BTX Division of Genetronics, Inc.

- Biolistics is a system for delivering polynucleotides into a target cell using microprojectile bombardment techniques. An illustrative embodiment of a method for delivering poly-
- 15 nucleotides into target cells by acceleration is a Biolistics Particle Delivery System, which can be used to propel particles coated with DNA or cells through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with cultured target cells. The screen disperses the particles so that they are not delivered to the target cells in large aggregates. It is believed that a screen intervening between the projectile apparatus and the cells to be
 - 20 bombarded reduces the size of projectiles aggregate and may contribute to a higher frequency of transformation by reducing damage inflicted on the recipient cells by projectiles that are too large.

- For the bombardment, cells in suspension are preferably concentrated on filters or solid culture medium. Alternatively, other target cells may be arranged on solid culture
- 25 medium. The cells to be bombarded are positioned at an appropriate distance below the microprojectile stopping plate. If desired, one or more screens are also positioned between the acceleration device and the cells to be bombarded. Through the use of these well-known techniques one may obtain up to 1000 or more foci of cells transiently expressing a marker gene. The number of cells in a focus which express the exogenous gene product 48
 - 30 hours post-bombardment often range from 1 to 10 and average 1 to 3.

In bombardment transformation, one may optimize the prebombardment culturing conditions and the bombardment parameters to yield the maximum numbers of stable transformants. Both the physical and biological parameters for bombardment are important in this technology. Physical factors are those that involve manipulating the polynucleotide/-

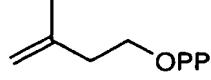
- 27 -

- microparticle precipitate or those that affect the flight and velocity of either the macro- or microparticles. Biological factors include all steps involved in manipulation of cells before and immediately after bombardment, the osmotic adjustment of target cells to help alleviate the trauma associated with bombardment, and also the nature of the transforming 5 DNA, such as linearized DNA or intact supercoiled plasmids.

- Accordingly, it is contemplated that one may wish to adjust various of the bombardment parameters in small-scale studies to fully optimize the conditions. One may particularly wish to adjust physical parameters such as gap distance, flight distance, tissue distance, and helium pressure. One may also minimize the trauma reduction factors (TRFs) by modifying 10 conditions which influence the physiological state of the recipient cells and which may therefore influence transformation and integration efficiencies. For example, the osmotic state, tissue hydration and the subculture stage or cell cycle of the recipient cells may be adjusted for optimum transformation. The execution of other routine adjustments will be known to those of skill in the art in light of the present disclosure.
- 15 The methods of particle-mediated transformation is well known to those of skill in the art. E.g. US 5,015,580 (specifically incorporated herein by reference) describes the transformation of soybeans using such a technique.

- Another embodiment of the invention is a method for engineering a bacterium to produce an isoprenoid compound. Such a bacterium is made by (a) culturing a parent bacterium 20 in a medium under conditions permitting expression of an isoprenoid, and selecting a mutant bacterium from the culture medium that produces about 1.1-1,000 times more of an isoprenoid than the parent bacteria; (b) introducing into the mutant bacterium an expression vector containing a polynucleotide sequence represented by SEQ ID NO:42 operably linked to an expression control sequence; and (c) selecting a bacterium that contains the expression vector and produces at least about 1.1 times more of an isoprenoid 25 than the mutant in step (a).

In this embodiment, an isoprenoid compound means a compound structurally based on isopentenyl diphosphate (IPP) units of the formula:



- 30 Such compounds include the hemiterpenes, monoterpenes, sesquiterpenes, diterpenes, triterpenes (e.g., phytosterols, phytoestrogens, phytoecdysones, estrogens, phytoestrogens),

- 28 -

tetraterpenes (carotenoids), and polyterpenes. Preferably, the isoprenoid is a carotenoid, such as for example, one of the carotenoids identified above, in particular zeaxanthin.

The bacterium may be any bacterium that is capable of producing an isoprenoid compound using the processes disclosed herein. Preferably, the bacterium is a *Paracoccus*,

- 5 *Flavobacterium*, *Agrobacterium*, *Alcaligenes*, *Erwinia*, *E. coli*, or *B. subtilis*. Even more preferably, the bacterium is a *Paracoccus*. Preferably, the parent bacterium is R-1506 or R 1512, and the mutant bacterium is R1534 or R114, preferably R114.

The bacterium is cultured in a media and under conditions that are optimized for the production of the isoprenoid. The selection of media and culture conditions are well within

- 10 the skill of the art. The assays set forth in Examples 1, 11, and 12 provide exemplary methods for measuring the presence of certain carotenoids in a culture media. By optimizing the culture conditions and measuring for the production of the target isoprenoid, the culturing and selection of a mutant that meets the specific production parameters recited herein may be met. In this way, a mutant bacterium producing from 15 about 1.1-1,000 times more of an isoprenoid than the parent bacterium may be selected. Preferably, the mutant bacterium produces from about 1.5-500 times more of an isoprenoid than the parent bacterium, such as for example, at least about 100 times or at least about 10 times more of an isoprenoid than the parent bacterium. That bacterium is then cultured and used in subsequent steps.

- 20 After selecting the mutant bacterium that produces the desired level of an isoprenoid, an expression vector is introduced into the bacterium using any of the methods set forth above or described in the examples. Any of the expression vectors defined herein may be introduced into the mutant cell. Preferably, the expression vector contains SEQ ID NO:42.

- Once the expression vector is introduced into the mutant bacteria, a stable transformant is 25 selected that produces at least about 1.1 times, such as about 5 to about 20 times, more of an isoprenoid than the untransformed mutant. The selected transformant is then cultured under conditions suitable for isoprenoid production, and then the isoprenoid is isolated from the cell or the culture medium.

- 30 A further step in this method is introducing a mutation into the mutant bacterium that results in the increased production of an isoprenoid compound by the bacterium. The mutation may be selected from at least one of the following: inactivating the poly-hydroxyalkanoate (PHA) pathway, increasing expression of acetyl-CoA acetyltransferase, increasing expression of FPP synthase, increasing expression of an enzyme in a carotenoid

- 29 -

biosynthetic pathway, and increasing the expression of an enzyme for converting isopentenyl diphosphate (IPP) to dimethylallyl diphosphate (DMAPP).

The inactivating of the PHA pathway may be achieved by selecting for a mutant bacterium that does not express a polypeptide encoded by *phaB* (nucleotide positions 1258-1980 of SEQ ID NO:177) or by disrupting expression of the wild type *phaB* gene by homologous recombination using SEQ ID NO:177 or fragments thereof.

In this method, increasing expression of acetyl-CoA acetyltransferase may be achieved by introducing into the mutant bacterium a vector containing a polynucleotide sequence represented by SEQ ID NO:175 or nucleotide positions 1-1170 of SEQ ID NO:177 operably linked to an expression control sequence. In this method, increasing expression of FPP synthase may be achieved by introducing into the mutant bacterium a vector containing a polynucleotide sequence represented by nucleotides 295-1158 of SEQ ID NO:157 operably linked to an expression control sequence. In this method, increasing expression of a carotenoid gene may be achieved by introducing into the mutant bacterium a vector comprising a polynucleotide sequence that encodes one or more enzymes in the carotenoid biosynthetic pathway, such as for example a polynucleotide sequence selected from the group consisting of SEQ ID NOs:180, 182, and 184 operably linked to an expression control sequence.

In this method, it is preferred that the isoprenoid compound is isopentenyl diphosphate (IPP). It is also preferred that the isoprenoid compound is a carotenoid, such as for example, phytoene, lycopene, β -carotene, zeaxanthin, canthaxanthin, astaxanthin, adonixanthin, cryptoxanthin, echinenone, adonirubin, and combinations thereof.

Another embodiment of the invention is a microorganism of the genus *Paracoccus*, which microorganism has the following characteristics: (a) a sequence similarity to SEQ ID NO:12 of >97% using a similarity matrix obtained from a homology calculation using GeneCompar v. 2.0 software with a gap penalty of 0%; (b) a homology to R-1512, R1534, R114 or R-1506 of >70% using DNA:DNA hybridization at 81.5°C; (c) a G+C content of its genomic DNA that varies less than 1% from the G+C content of the genomic DNA of R114, R-1512, R1534, and R-1506; and (d) an average DNA fingerprint that clusters at about 58% similarity to strains R-1512, R1534, R114 and R-1506 using the AFLP procedure of Example 2, with the proviso that the microorganism is not *Paracoccus* sp. (MBIC3966).

- 30 -

Methods for determining each of these characteristics are fully set forth in Example 2, and it is contemplated when these methods are used that microorganisms meeting the above criteria will be readily identifiable. It is preferred that a microorganism of the present invention have each characteristic set forth above (i.e., a-d). However, any combination of 5 the characteristics a-d, which provides sufficient information to taxonomically validly describe a microorganism belonging to the same species as R114, R-1512, R1534, and R-1506, with the exception of *Paracoccus* sp. (MBIC3966) is also within the scope of the invention.

Another embodiment of the invention is a microorganism of the genus *Paracoccus*, which 10 microorganism has the following characteristics: (a) 18:1w7c comprising at least about 75% of the total fatty acids of the cell membranes; (b) an inability to use adonitol, i-erythritol, gentiobiose, β -methylglucoside, D-sorbitol, xylitol and quinic acid as carbon sources for growth; and (c) an ability to use L-asparagine and L-aspartic acid as carbon sources for growth, with the proviso that the microorganism is not *Paracoccus* sp. 15 (MBIC3966).

Methods for determining each of these characteristics are also fully set forth in Example 2, and it is contemplated when these methods are used that microorganisms meeting the above criteria will be readily identifiable. It is preferred that a microorganism of the present invention have each characteristic set forth above (i.e., a-c). However, any combination of 20 the characteristics a-c, which provides sufficient information to taxonomically validly describe a microorganism belonging to the same species as R114, R-1512, R1534, and R-1506, with the exception of *Paracoccus* sp. (MBIC3966) is also within the scope of the invention.

Another embodiment of the invention is a microorganism of the genus *Paracoccus*, which 25 microorganism has the following characteristics: (a) an ability to grow at 40°C; (b) an ability to grow in a medium having 8% NaCl; (c) an ability to grow in a medium having a pH of 9.1; and (d) a yellow-orange colony pigmentation, with the proviso that the microorganism is not *Paracoccus* sp. (MBIC3966).

Methods for determining each of these characteristics are also fully set forth in Example 2, 30 and it is contemplated when these methods are used that microorganisms meeting the above criteria will be readily identifiable. It is preferred that a microorganism of the present invention have each characteristic set forth above (i.e., a-d). However, any combination of the characteristics a-d, which provides sufficient information to taxonomically validly describe a microorganism belonging to the same species as R114, R-1512, R1534,

and R-1506, with the exception of *Paracoccus* sp. (MBIC3966) is also within the scope of the invention.

A microorganism of the present invention may also be identified using any combination of the 11 characteristics set forth above, which provide sufficient information to taxonomically validly describe a microorganism belonging to the same species as R114, R-1512, 5 R1534, and R-1506, with the exception of *Paracoccus* sp. (MBIC3966).

In accordance with the foregoing the present invention provides

- (1) an isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown as residues 1 to 340 of SEQ ID NO:43, in particular an amino acid sequence corresponding to positions 68-97 of SEQ ID NO:43;
 - (b) an amino acid sequence shown as residues 1 to 349 of SEQ ID NO:45, in particular an amino acid sequence corresponding to positions 1-30 of SEQ ID NO:45;
 - (c) an amino acid sequence shown as residues 1 to 388 of SEQ ID NO:47, in particular 10 an amino acid sequence corresponding to positions 269-298 of SEQ ID NO:47;
 - (d) an amino acid sequence shown as residues 1 to 378 of SEQ ID NO:49, in particular an amino acid sequence corresponding to positions 109-138 of SEQ ID NO:49;
 - (e) an amino acid sequence shown as residues 1 to 305 of SEQ ID NO:51, in particular 15 an amino acid sequence corresponding to positions 198-227 of SEQ ID NO:51;
 - (f) an amino acid sequence shown as residues 1 to 332 of SEQ ID NO:53, in particular an amino acid sequence corresponding to positions 81-110 of SEQ ID NO:53;
 - (g) a fragment of an amino acid sequence selected from the group consisting of SEQ 20 ID NOS: 43, 45, 47, 49, 51, and 53, wherein said fragment has at least 30 contiguous amino acid residues;
 - (h) an amino acid sequence of a fragment of a polypeptide selected from the group 25 consisting of SEQ ID NOS: 43, 45, 47, 49, 51, and 53, the fragment having the activity of hydroxymethylglutaryl-CoA reductase (HMG-CoA reductase), isopentenyl diphosphate isomerase, hydroxymethylglutaryl-CoA synthase (HMG-CoA synthase), mevalonate kinase, phosphomevalonate kinase, or diphosphomevalonate decarboxylase;
 - (i) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe comprising at least 30 consecutive nucleotides of SEQ ID NO:42 or a complement of SEQ ID NO:42, wherein 30 the polypeptide has the activity of HMG-CoA reductase, isopentenyl diphosphate

- 32 -

isomerase, HMG-CoA synthase, mevalonate kinase, phosphomevalonate kinase, or diphosphomevalonate decarboxylase; and

(j) a conservatively modified variant of SEQ ID NO:43, 45, 47, 49, 51 or 53.

(2) an isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

5 (a) an amino acid sequence shown as residues 1 to 287 of SEQ ID NO:159;

(b) at least 30 contiguous amino acid residues of SEQ ID NO:159;

(c) an amino acid sequence of a fragment of SEQ ID NO: 159, the fragment having the activity of farnesyl diphosphate synthase (FPP synthase);

10 (d) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe comprising at least 30 consecutive nucleotides spanning positions 295-1158 of SEQ ID NO:157 or a complement thereof, wherein the polypeptide has the activity of FPP synthase; and

(e) a conservatively modified variant of SEQ ID NO:159.

15 (3) an isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence shown as residues 1 to 142 of SEQ ID NO:160;

(b) at least 30 contiguous amino acid residues of SEQ ID NO:160;

20 (c) an amino acid sequence of a fragment of SEQ ID NO: 160, the fragment having the activity of 1-deoxyxylulose-5-phosphate synthase (DXPS);

(d) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe comprising at least 30 consecutive nucleotides spanning positions 1185-1610 of SEQ ID NO:157 or a complement thereof, wherein the polypeptide has the activity of DXPS;

25 (e) a conservatively modified variant of SEQ ID NO:160.

(4) an isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence shown as residues 1 to 390 of SEQ ID NO:178;

(b) at least 30 contiguous amino acid residues of SEQ ID NO:178;

30 (c) an amino acid sequence of a fragment of a polypeptide of SEQ ID NO: 178, the fragment having the activity of acetyl-CoA acetyltransferase; -

(d) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe comprising at least 30 consecutive nucleotides spanning positions 1-1170 of SEQ ID NO:177 or a complement thereof, wherein the polypeptide has the activity of acetyl-CoA acetyltransferase; and

35 (e) a conservatively modified variant of SEQ ID NO:178.

(5) an isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

- (a) an amino acid sequence shown as residues 1 to 240 of SEQ ID NO:179;
- (b) at least 30 contiguous amino acid residues of SEQ ID NO:179;
- 5 (c) an amino acid sequence of a fragment of a polypeptide of SEQ ID NO: 179, the fragment having the activity of acetoacetyl-CoA reductase;
- (d) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe comprising at least 30 consecutive nucleotides spanning positions 1258-1980 of SEQ ID NO:177 or a complement thereof, wherein the polypeptide has the activity of acetoacetyl-CoA reductase;
- 10 and
- (e) a conservatively modified variant of SEQ ID NO:179.

(6) an isolated polynucleotide sequence comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:42, variants of SEQ ID NO:42 containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of SEQ ID NO:42 that encode a polypeptide having an activity selected from the group consisting of hydroxymethylglutaryl-CoA reductase (HMG-CoA reductase), isopentenyl diphosphate isomerase, hydroxymethylglutaryl-CoA synthase (HMG-CoA synthase), mevalonate kinase, phosphomevalonate kinase, and diphosphomevalonate decarboxylase, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides of SEQ ID NO:42, or the complement of SEQ ID NO:42, which polynucleotide encodes a polypeptide having an activity selected from the group consisting of HMG-CoA reductase, isopentenyl diphosphate isomerase, HMG-CoA synthase, mevalonate kinase, phosphomevalonate kinase, and diphosphomevalonate decarboxylase; in particular

- (a) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of nucleotides 2622 to 3644 of SEQ ID NO:42, fragments thereof that encode a polypeptide having HMG-CoA reductase activity, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning residues 2622 to 3644 of SEQ ID NO:42, or a complement thereof, wherein the polynucleotide encodes a polypeptide having HMG-CoA reductase activity, more particularly a polynucleotide sequence consisting of nucleotides 2622 to 3644 of SEQ ID NO:42;
- 30 (b) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of nucleotides 3641 to 4690 of SEQ ID NO:42,

- 34 -

- variants thereof containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of SEQ ID NO:42 that encode a polypeptide having isopentenyl diphosphate isomerase activity, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning residues 3641 to 4690 of SEQ ID NO:42, or a complement thereof, wherein the polynucleotide encodes a polypeptide having isopentenyl diphosphate isomerase activity, more particularly a polynucleotide sequence consisting of nucleotides 3641 to 4690 of SEQ ID NO:42.
- 5 (c) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of nucleotides 4687 to 5853 of SEQ ID NO:42, variants thereof containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of SEQ ID NO:42 that encode a polypeptide having HMG-CoA synthase activity, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning residues 4687 to 5853 of SEQ ID NO:42, or a complement thereof, wherein the polynucleotide encodes a polypeptide having HMG-CoA synthase activity, more particularly a polynucleotide sequence consisting of nucleotides 3641 to 4690 of SEQ ID NO:42;
- 10 (d) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of nucleotides 5834 to 6970 of SEQ ID NO:42, variants thereof containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of SEQ ID NO:42 that encode a polypeptide having mevalonate kinase activity, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning residues 5834 to 6970 of SEQ ID NO:42, or a complement thereof, wherein the polynucleotide encodes a polypeptide having mevalonate kinase activity, more particularly a polynucleotide sequence consisting of nucleotides 3641 to 4690 of SEQ ID NO:42;
- 15 (e) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of nucleotides 6970 to 7887 of SEQ ID NO:42, variants thereof containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of SEQ ID NO:42 that encode a polypeptide having phosphomevalonate kinase activity, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning residues 6970 to 7887 of SEQ ID NO:42, or a complement thereof, wherein the polynucleotide encodes a
- 20
- 25
- 30
- 35

- 35 -

- polypeptide having phosphomevalonate kinase activity, more particularly a polynucleotide sequence consisting of nucleotides 3641 to 4690 of SEQ ID NO:42; or
5 (f) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of nucleotides 7880 to 8878 of SEQ ID NO:42, variants thereof containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of SEQ ID NO:42 that encode a polypeptide having diphosphomevalonate decarboxylase activity, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning residues 7880 to 8878 of SEQ ID NO:42, or a complement thereof, wherein the polynucleotide encodes a polypeptide having diphosphomevalonate decarboxylase activity, more particularly an isolated polynucleotide consisting of nucleotides 7880 to 8878 of SEQ ID NO:42, more particularly a polynucleotide sequence consisting of nucleotides 3641 to 4690 of SEQ ID NO:42;
- 10 15 (7) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of the nucleotide sequence of SEQ ID NO:157, variants of SEQ ID NO:157 containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of SEQ ID NO:157 that encode a polypeptide having farnesyl diphosphate (FPP) synthase activity, 1-deoxy-D-xylulose 5-phosphate synthase activity or a polypeptide having the activity of XseB, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides of SEQ ID NO:157, or the complement of SEQ ID NO:157, which polynucleotide encodes a polypeptide having an activity selected from the 20 25 group consisting of FPP synthase activity, 1-deoxy-D-xylulose 5-phosphate synthase activity, and the activity of XseB, in particular
(a) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of a nucleotide sequence spanning positions 59-292 of SEQ ID NO:157, variants thereof containing one or more substitutions according to the 30 35 *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of the nucleotide sequence spanning positions 59-292 of SEQ ID NO:157 that encode a polypeptide having the function of XseB, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning positions 59-292 of SEQ ID NO:157, or the complement of such a sequence, wherein the polynucleotide encodes a polypeptide having the function of XseB, more particularly an isolated polynucleotide consisting of nucleotides 59 to 292 of SEQ ID NO:157;

- (b) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of the nucleotide sequence spanning positions 295-1158 of SEQ ID NO:157, variants of the nucleotide sequence spanning positions 295-1158 of SEQ ID NO:157 containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of the nucleotide sequence spanning positions 295-1158 of SEQ ID NO:157 that encode a FPP synthase activity, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning positions 295-1158 of SEQ ID NO:157, or the complement of such a sequence, wherein the polynucleotide encodes a polypeptide having FPP synthase activity, more particularly an isolated polynucleotide consisting of nucleotides 295 to 1158 of SEQ ID NO:157;
- (c) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of the nucleotide sequence spanning positions 1185-1610 of SEQ ID NO:157, variants of the nucleotide sequence spanning positions 1185-1610 of SEQ ID NO:157 containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of the nucleotide sequence spanning positions 1185-1610 of SEQ ID NO:157 that encode a polypeptide having 1-deoxyxylulose-5-phosphate synthase activity, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning positions 1185-1610 of SEQ ID NO:157, or the complement of such a sequence, wherein the polynucleotide encodes a polypeptide having 1-deoxyxylulose-5-phosphate synthase activity, more particularly an isolated polynucleotide consisting of nucleotides 1185 to 1610 of SEQ ID NO:157;
- (8) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of the nucleotide sequence of SEQ ID NO:177, variants of SEQ ID NO:177 containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of SEQ ID NO:177 that encode a polypeptide having an activity selected from the group consisting of acetyl-CoA acetyltransferase and acetoacetyl-CoA reductase, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides of SEQ ID NO:177, or the complement of SEQ ID NO:177, which polynucleotide encodes a polypeptide having an activity selected from the group consisting of acetyl-CoA acetyltransferase and acetoacetyl-CoA reductase, in particular

- 37 -

- (a) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of nucleotides 1 to 1170 of SEQ ID NO:177, variants of SEQ ID NO:177 containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of SEQ ID NO:177 that encode a polypeptide having acetyl-CoA acetyltransferase activity, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning residues 1 to 1170 of SEQ ID NO:177, or a complement thereof, wherein the polynucleotide encodes a polypeptide having acetyl-CoA acetyltransferase activity, more particularly an isolated polynucleotide sequence consisting of nucleotides 1-1170 of SEQ ID NO:177;
- (b) an isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of nucleotides 1258-1980 of SEQ ID NO:177, variants of SEQ ID NO:177 containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table (Table 14), fragments of SEQ ID NO:177 that encode a polypeptide having acetoacetyl-CoA reductase activity, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides spanning residues 1258-1980 of SEQ ID NO:177, or a complement thereof, wherein the polynucleotide encodes a polypeptide having acetoacetyl-CoA reductase activity, more particularly an isolated polynucleotide sequence consisting of nucleotides 1258-1980 of SEQ ID NO:177;
- (9) an isolated polynucleotide sequence comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:42, SEQ ID NO:157, SEQ ID NO:177, and combinations thereof;
- (10) an expression vector comprising the polynucleotide sequence of any one of (6) (a) to (6) (f), (7) (a) to (7) (c), (8) (a), (8) (b) or (9), in particular an expression vector wherein the polynucleotide sequence is operably linked to an expression control sequence, e.g. an expression vector further comprising a polynucleotide sequence that encodes an enzyme in the carotenoid biosynthetic pathway, more particularly an expression vector wherein the polynucleotide sequence is selected from the group consisting of SEQ ID NO:180, SEQ ID NO:182, SEQ ID NO:184, and combinations thereof which are operably linked to an expression control sequence;
- (11) an expression vector selected from the group consisting of pBBR-K-mev-op16-1, pBBR-K-mev-op16-2, pDS-*mvaA*, pDS-*idi*, pDS-*hcs*, pDS-*mvk*, pDS-*pmk*, pDS-*mvd*, pDS-*His-mvaA*, pDS-*His-idi*, pDS-*His-hcs*, pDS-*His-mvk*, pDS-*His-pmk*, pDS-*His-mvd*, pBBR-K-Zea4, pBBR-K-Zea4-up, pBBR-K-Zea4-down, pBBR-K-PcrtE-crtE-3,

- pBBR-tK-PcrtE-mvaA, pBBR-tK-PcrtE-idi, pBBR-tK-PcrtE-hcs, pBBR-tK-PcrtE-mvk,
pBBR-tK-PcrtE-pmk, pBBR-tK-PcrtE-mvd, pBBR-K-PcrtE-mvaA-crtE-3, pDS-His-
phaA, pBBR-K-PcrtE-crtW, pBBR-K-PcrtE-crtWZ, pBBR-K-PcrtE-crtZW, and
combinations thereof, in particular
- 5 (a) an expression vector selected from the group consisting of pBBR-K-mev-op16-1
and pBBR-K-mev-op16-2,
- (b) an expression vector selected from the group consisting of pBBR-K-Zea4, pBBR-
K-Zea4-up, and pBBR-K-Zea4-down;
- 10 (c) an expression vector selected from the group consisting of pBBR-K-PcrtE-crtE-3,
pBBR-tK-PcrtE-mvaA, pBBR-tK-PcrtE-idi, pBBR-tK-PcrtE-hcs, pBBR-tK-PcrtE-mvk,
pBBR-tK-PcrtE-pmk, pBBR-tK-PcrtE-mvd, and combinations thereof;
- (d) an expression vector which is pBBR-K-PcrtE-mvaA-crtE-3;
- 15 (e) an expression vector which is pDS-His-phaA; or
- (f) an expression vector selected from the group consisting of pBBR-K-PcrtE-crtW,
pBBR-K-PcrtE-crtWZ, and pBBR-K-PcrtE-crtZW;
- (12) a cultured cell comprising the polynucleotide sequence of any one of (6) (a) to (f), (7)
20 (a) to (c), (8) (a), (8) (b) or (9), or an expression vector of (10) or (11), or a progeny
of the cell, wherein the cell expresses a polypeptide encoded by the polynucleotide
sequence, in particular a cell which is further characterized by a feature selected from
(a) further comprising a polynucleotide sequence that encodes an enzyme in the
carotenoid biosynthetic pathway, more particularly a cultured cell wherein the poly-
nucleotide sequence that encodes an enzyme in the carotenoid biosynthetic pathway is
selected from the group consisting of SEQ ID NOs:180, 182, and 184, or a progeny of
the cell, wherein the cell expresses polypeptides encoded by the polynucleotide
25 sequences, and
- (b) from being a member of a group selected from yeast, fungus, bacterium and alga,
in particular a bacterium selected from the group consisting of *Paracoccus*, *Flavo-*
bacterium, *Agrobacterium*, *Alcaligenes*, *Erwinia*, *E. coli*, and *B. subtilis*, more particularly
Paracoccus, more particularly *Paracoccus* selected from the group consisting of R-1506,
30 R-1512, R1534, and R114;
- (13) a method of producing a carotenoid comprising culturing a cell of (12) under condi-
tions permitting expression of a polypeptide encoded by the polynucleotide sequence,
and isolating the carotenoid from the cell or the medium of the cell;
- (14) a method of making a carotenoid-producing cell comprising:
- 35 (a) introducing into a cell a polynucleotide sequence encoding an enzyme in the
mevalonate pathway, which enzyme is expressed in the cell; and

- 39 -

- (b) selecting a cell containing the polynucleotide sequence of step (a) that produces a carotenoid at a level that is about 1.1-1,000 times the level of the carotenoid produced by the cell before introduction of the polynucleotide sequence,
in particular a method selected from a method characterized by a feature selected
5 from
- (i) the selecting step comprising selecting a cell containing the polynucleotide sequence of step (a) that produces a carotenoid at a level that is about 1.5-500 times, particularly about 100 times, or at least about 10 times, the level of the carotenoid produced by the cell before introduction of the polynucleotide sequence;
- 10 (ii) the cell producing from about 1 mg/L to about 10 g/L of a carotenoid.
- (iii) the cell being selected from the group consisting of a yeast, fungus, bacterium, and alga, in particular selected from the group consisting of *Paracoccus*, *Flavobacterium*, *Agrobacterium*, *Alcaligenes*, *Erwinia*, *E. coli*, and *B. subtilis*, more particularly from *Paracoccus*;
- 15 (iv) the cell in step (a) being a mutant cell, in particular being selected from the group consisting of R114 and R1534, in particular the mutant cell producing about 1.1-1,000 times, in particular about 1.5-500 times, more particularly at least about 100 times more or at least about 10 times more, the level of a carotenoid compared to its non-mutant parent;
- 20 (v) the polynucleotide sequence being selected from polynucleotide sequences of (6) (a) to (f), (7) (a) to (c), (8) (a), (8) (b) and (9), in particular wherein the polynucleotide sequence is operably linked to an expression control sequence;
- (vi) the polynucleotide sequence being an expression vector of (10) or (11);
- (vii) the introducing step being selected from the group consisting of transformation, transduction, transfection, lipofection, electroporation, conjugation, and biolistics.
25 (viii) the carotenoid being selected from the group consisting of phytoene, lycopene, β-carotene, zeaxanthin, canthaxanthin, astaxanthin, adonixanthin, cryptoxanthin, echinenone, adonirubin, and combinations thereof, in particular the carotenoid being zeaxanthin;
- 30 (15) a method for engineering a bacterium to produce an isoprenoid compound comprising:
- (a) culturing a parent bacterium in a medium under conditions permitting expression of an isoprenoid compound, and selecting a mutant bacterium from the culture medium that produces about 1.1-1,000 times more of an isoprenoid compound than the parent bacterium;
35

- 40 -

- (b) introducing into the mutant bacterium an expression vector comprising a polynucleotide sequence represented by SEQ ID NO:42 operably linked to an expression control sequence; and
- 5 (c) selecting a bacterium that contains the expression vector and produces at least about 1.1 times more of an isoprenoid compound than the mutant in step (a), in particular
- (i) a method further comprising introducing a mutation into the mutant bacterium, more particularly a method wherein the mutation causes an effect selected from at least one of the following: inactivating the polyhydroxyalkanoate (PHA) pathway, increasing expression of acetyl-CoA acetyltransferase, increasing expression of farnesyl diphosphate (FPP) synthase, increasing expression of an enzyme in a carotenoid pathway, increasing the expression of an enzyme for converting isopentenyl diphosphate (IPP) to dimethylallyl diphosphate (DMAPP),
- 10 most particularly a method wherein inactivating of the PHA pathway comprises selecting for a mutant bacterium that does not express a polypeptide encoded by *phab* (nucleotide positions 1258-1980 of SEQ ID NO: 177) or by disrupting expression of the wild type *phab* gene by homologous recombination using SEQ ID NO:177 or a fragment thereof, or
- 15 a method wherein increasing expression of acetyl-CoA acetyltransferase comprises introducing into the mutant bacterium a vector comprising a polynucleotide sequence represented by SEQ ID NO:175 or nucleotide positions 1-1170 of SEQ ID NO:177 operably linked to an expression control sequence, or
- 20 a method wherein increasing expression of FPP synthase comprises introducing into the mutant bacterium a vector comprising a polynucleotide sequence represented by nucleotides 295-1158 of SEQ ID NO:157 operably linked to an expression control sequence, or
- 25 a method wherein increasing expression of an enzyme in a carotenoid pathway comprises introducing into the mutant bacterium a vector comprising a polynucleotide sequence selected from the group consisting SEQ ID NOs:180, 182, and 184 operably linked to an expression control sequence;
- 30 (b) a method wherein the isoprenoid is isopentenyl diphosphate (IPP).
- (c) a method wherein the isoprenoid is a carotenoid, in particular a method wherein the carotenoid is selected from the group consisting of phytoene, lycopene, β -carotene, zeaxanthin, canthaxanthin, astaxanthin, adonixanthin, cryptoxanthin, echinenone, adonirubin, and combinations thereof;
- 35

- 41 -

- (d) a method wherein the parent bacterium is a *Paracoccus*, in particular R-1512 or R-1506, or R1534 or R114, in particular wherein the mutant is R114;
- (16) a microorganism of the genus *Paracoccus*, which microorganism has the following characteristics:
- 5 (i) a sequence similarity to SEQ ID NO:12 of >97% using a similarity matrix obtained from a homology calculation using GeneCompar v. 2.0 software with a gap penalty of 0%;
a homology to strain R-1512, R1534, R114 or R-1506 of >70% using DNA:DNA hybridization at 81.5°C;
- 10 a G+C content of its genomic DNA that varies less than 1% from the G+C content of the genomic DNA of R114, R-1512, R1534, and R-1506; and
an average DNA fingerprint that clusters at about 58% similarity to strains R-1512, R1534, R114 and R-1506 using the AFLP procedure of Example 2, with the proviso that the microorganism is not *Paracoccus* sp. (MBIC3966);
- 15 (ii) 18:1w7c comprising at least about 75% of the total fatty acids of the cell membranes;
an inability to use adonitol, i-erythritol, gentiobiose, β-methylglucoside, D-sorbitol, xylitol and quinic acid as carbon sources for growth; and
an ability to use L-asparagine and L-aspartic acid as carbon sources for growth, with
20 the proviso that the microorganism is not *Paracoccus* sp. (MBIC3966); or
(iii) an ability to grow at 40°C;
an ability to grow in a medium having 8% NaCl;
an ability to grow in a medium having a pH of 9.1; and
a yellow-orange colony pigmentation, with the proviso that the microorganism is not
25 *Paracoccus* sp. (MBIC3966).

The following examples are provided to further illustrate certain aspects of the present invention. These examples are illustrative only and are not intended to limit the scope of the invention in any way.

Example 1: Analytical and Biochemical Methods

30 (a) Analysis of Carotenoids

Sample preparation. A solvent mixture of 1:1 dimethylsulfoxide (DMSO) and tetrahydrofuran (THF) was first prepared. This solvent mixture was stabilized by the addition of butylated hydroxytoluene (BHT, 0.5 g/l solvent mixture). Four milliliters of the stabilized DMSO/THF mixture was added to 0.4 ml of bacterial culture in a disposable 15-ml poly-

- 42 -

propylene centrifuge tube (gives a final dilution factor of 1/11). The tubes were capped and mixed using a Vortex mixer for 10 seconds each. The samples were then put on a Brinkmann Vibramix shaker for 20 minutes. The tubes were centrifuged at room temperature for 4 minutes at 4000 rpm and aliquots of the clear yellow/orange supernatant 5 were transferred into brown glass vials for analysis by High Performance Liquid Chromatography (HPLC).

HPLC. A reversed phase HPLC method was developed for the simultaneous determination of astaxanthin, zeaxanthin, canthaxanthin, β -carotene, and lycopene. The method was also able to separate the main *cis*-isomers of zeaxanthin. Chromatography was performed using an Agilent 1100 HPLC system equipped with a thermostatted autosampler and a diode array detector. The method parameters were as follows:

Column: YMC Carotenoid C30 column, particle size 5 micron

250* 4.6mm I.D., steel

(YMC, Part No. CT99S052546WT)

15 Guard column: Pelliguard LC-18 cartridge, 20 mm
(SUPELCO, Part No. 59654)

Mobile phase: Methanol (MeOH)/Methyl tert-butyl ether (TBME) gradient

	% MeOH	% TBME
Start	80	20
10 min	65	35
20 min	10	90

Run time: 28 min; Typical column pressure: 90 bar at start; Flow rate: 1.0 ml/min.;

20 Detection: UV at 450 nm; Injection volume: 10 μ l; Column temperature: 15°C

Reagents. Methanol and TBME were HPLC grade and were obtained from EM Science and J.T. Baker, respectively. DMSO (Omnisolve) was purchased from EM Science. THF (HPLC solvent) was from Burdick and Jackson.

Calculations. Quantitative analyses were performed with a two level calibration using 25 external standards (provided by Hoffmann-La Roche, Basel, Switzerland). Calculations were based on peak areas.

Selectivity. The selectivity of the method were verified by injecting standard solutions of the relevant carotenoid reference compounds. The target compounds (all-*trans*-carotenoids) were completely separated and showed no interference. Some minor *cis* isomers may 30 coelute, although these potentially interfering isomers are rare and need not be considered in routine analyses. The retention times of the compounds are listed in Table 1.

Table 1. HPLC retention times for carotenoids.

Carotenoid	Retention time (min.)	Carotenoid	Retention time (min.)
Astaxanthin	6.99	Canthaxanthin	9.95
Adonixanthin	7.50	Cryptoxanthin	13.45
15-cis-Zeaxanthin	7.80	β-Carotene	17.40
13-cis-Zeaxanthin	8.23	Lycopene	21.75
all-trans-Zeaxanthin	9.11		

Linearity. 25 Milligrams of all-trans-zeaxanthin were dissolved in 50 ml of DMSO/THF mixture (giving a final zeaxanthin concentration 500 µg/ml). A dilution series was prepared (final zeaxanthin concentrations of 250, 100, 50, 10, 5, 1, and 0.1 µg/ml) and analyzed by the HPLC method described above. A linear range was found from 0.1 µg/ml to 250 µg/ml. The correlation coefficient was 0.9998.

Limit of detection. The lower limit of detection for zeaxanthin by this method was determined to be 60 µg/l. A higher injection volume and optimization of the integration parameters made it possible to lower the detection limit to approximately 5 µg/l.

Reproducibility. The retention time for all-trans-zeaxanthin was very stable (relative standard deviation (RSD), 0.2 %). The peak area reproducibility, based on ten repetitive analyses of the same culture sample, was determined to be 0.17 % RSD for all trans-zeaxanthin and 1.0 % for cryptoxanthin.

15 (b) Preparation of crude extracts and enzyme assay methods.

Preparation of crude extracts. Crude extracts of *Paracoccus* and *E. coli* were prepared by resuspending washed cell pellets in 1 ml of extraction buffer (buffer used depended on the enzyme being assayed – compositions are specified along with each enzyme assay procedure described below). Cell suspensions were placed in a 2-ml plastic vial and disrupted by agitation with glass beads using a Mini Bead Beater 8 (Biospec Products, Bartlesville, OK, USA). Disruption was performed at 4°C using a medium agitation setting. The disrupted preparations were centrifuged at 21,000 x g for 20 minutes at 4°C to sediment the cell debris, and the supernatants were used directly for enzyme assays.

Protein determinations. Protein concentrations in crude extracts were determined by the method of Bradford [Anal. Biochem. 72:248-254 (1976)] using the Bio-Rad Protein Assay

- 44 -

Reagent (Bio-Rad, Hercules, CA, USA). Bovine serum albumin was used as the reference protein for construction of standard curves.

Acetyl-CoA acetyltransferase assays. Crude extracts were prepared in 150 mM EPPS (*N*-[2-hydroxyethyl] piperazine-*N'*-[3-propanesulfonic acid]) buffer, pH 8.0. Assays were 5 performed in the thiolysis direction according to the method described by Slater et al. [J. Bacteriol. 180:1979-1987 (1998)]. This assay measures the disappearance of acetoacetyl-CoA spectrophotometrically at 304 nm. Reaction mixtures contained 150 mM EPPS buffer (pH 8.0), 50 mM MgCl₂, 100 μM CoA, 40 μM acetoacetyl-CoA and crude extract. Reactions were carried out at 30°C and were initiated by addition of crude extract. The 10 disappearance of acetoacetyl-CoA at 304 nm was monitored using a SpectraMAX Plus plate reader (Molecular Devices Corp., Sunnyvale, CA, USA) and a quartz microtiter plate (any standard spectrophotometer can also be used). Activity (expressed as U/mg protein) was calculated using a standard curve constructed with acetoacetyl-CoA (1 unit of activity = 1 μmol acetoacetyl-CoA consumed/min.). The lower limit of detection of Acetyl-CoA 15 acetyltransferase activity was 0.006 U/mg.

HMG-CoA synthase assays. HMG-CoA synthase was assayed according to the method of Honda et al. [Hepatology 27:154-159 (1998)]. In this assay, the formation of HMG-CoA from acetyl-CoA and acetoacetyl-CoA is measured directly by separating the reaction product and substrates by HPLC. Crude extracts were prepared in 50 mM Tris-HCl buffer 20 (pH 8.0). Reaction mixtures (0.1 ml) contained 50 mM Tris-HCl buffer (pH 8.0), 0.1 mM EDTA, 20 mM MgCl₂, 0.1 mM acetoacetyl-CoA, 0.8 mM acetyl-CoA and crude extract. Reactions were pre-incubated for 2 minutes at 30°C before adding the crude extract. After 25 5 minutes of reaction at 30°C, the reactions were stopped by adding 0.2 ml of 200 mM tetra-butyl ammonium phosphate (TBAP, dissolved in methanol-water (3:2, final pH was 5.5) and containing 0.2 mM propionyl-CoA as an internal recovery standard). The mixture was then centrifuged for 3 minutes at 21,000 x g at 4°C and subsequently kept on ice until analyzed by reversed phase ion-pair HPLC. HMG-CoA and propionyl-CoA were 30 separated from acetyl-CoA and acetoacetyl-CoA using a Nova-Pak C18 column (3.9 x 150 mm, Waters Corporation, Milford, MA, USA). The injection volume was 20 μl, the mobile phase was 50 mM TBAP dissolved in methanol-water (1:1, final pH was 5.5), and the flow rate was 1.0 ml/min. HMG-CoA and propionyl-CoA were detected by absorbance at 254 nm. HMG-CoA produced in the reaction was quantified by comparison with a standard curve created using authentic HMG-CoA. Activity is defined as U/mg protein. One unit of activity = 1 nmol HMG-CoA produced/min. The lower 35 limit of detection of HMG-CoA synthase was about 1 U/mg.

- 45 -

HMG-CoA reductase assays. Crude extracts were prepared in 25 mM potassium phosphate buffer (pH 7.2) containing 50 mM KCl, 1 mM EDTA and a protease inhibitor cocktail (Sigma Chemical Co., St. Louis, MO, USA, catalog #P-2714). Assays were performed according to the method of Takahashi et al. [J. Bacteriol. 181:1256-1263 (1999)]. This assay measures the HMG-CoA dependent oxidation of NADPH spectrophotometrically at 340nm. Reaction mixtures contained 25 mM potassium phosphate buffer (pH 7.2), 50 mM KCl, 1 mM EDTA, 5 mM dithiothreitol, 0.3 mM NADPH, 0.3 mM R,S-HMG-CoA and crude extract. Reactions were performed at 30°C and were initiated by the addition of HMG-CoA. HMG-CoA-dependent oxidation of NADPH was monitored at 340 nm using a SpectraMAX Plus plate reader (Molecular Devices Corp., Sunnyvale, CA, USA) and a quartz microtiter plate (any standard spectrophotometer may be used). Activity (expressed as U/mg protein) was calculated using a standard curve constructed with NADPH (1 unit of activity = 1 µmol NADPH oxidized/min.). The lower limit of detection of HMG-CoA reductase activity was 0.03 U/mg.

Mevalonate kinase, phosphomevalonate kinase and mevalonate diphosphate decarboxylase assays. The preparation of substrates and the assay procedures for mevalonate kinase, phosphomevalonate kinase and mevalonate diphosphate decarboxylase have been described in detail by Popják [Methods Enzymol. 15:393-425 (1969)]. For all assays, one unit of enzyme activity is defined as 1µmol of product formed/minute. In addition to these spectrophotometric and radiochromatographic assays, alternate methods, for example using HPLC separation of reaction substrates and products, can be used. The lower limit of detection of mevalonate kinase, phosphomevalonate kinase and mevalonate diphosphate decarboxylase is typically about 0.001 U/mg protein.

IPP isomerase assays. Crude extracts were prepared in 50 mM Tris-HCl buffer (pH 7.5). Assays were performed using the method of Spurgeon et al. [Arch. Biochem. Biophys. 230:445-454 (1984)]. This assay is based on the difference in acid-lability of IPP and DMAPP. Reaction mixtures (0.1 ml final volume) contained 50 mM Tris-HCl buffer (pH 7.5), 2 mM dithiothreitol, 5 mM MgCl₂, 20 µM [1-¹⁴C]-IPP and crude extract. Reactions were carried out at 30°C for 15 minutes and terminated by the addition of 0.3 ml of a mixture of concentrated HCl:methanol (4:1) and an additional incubation at 37°C for 20 min. Hexane (0.9 ml) was added and the tubes were mixed (4 times for 10 seconds using a vortex mixer). After centrifugation (21,000 x g, 5 minutes), 0.6 ml of the hexane layer was transferred to a scintillation vial, scintillation fluid was added, and the radioactivity counted. Activity is expressed as U/mg protein. One unit of activity = 1 pmol [1-¹⁴C]-IPP incorporated into acid labile products/min. The lower limit of detection of IPP isomerase activity was 1 U/mg.

- 46 -

FPP synthase assays. Crude extracts were prepared in 50 mM Tris-HCl buffer (pH 8.0). The FPP synthase assay procedure was similar to the IPP isomerase assay described above, being based on the difference in acid lability of IPP and FPP (Spurgeon et al., *supra*). Reaction mixtures (0.1 ml final volume) contained 50 mM Tris-HCl buffer (pH 8.0), 5 2 mM dithiothreitol, 5 mM MgCl₂, 20 µM [1-¹⁴C]-IPP, 25 µM GPP (geranyl pyrophosphate) and crude extract. Reactions were carried out at 30°C for 15 minutes and terminated by the addition of 0.3 ml of a mixture of concentrated HCl:methanol (4:1) and an additional incubation at 37°C for 20 minutes. Hexane (0.9 ml) was added and the tubes were mixed (4x, 10 seconds using a vortex mixer). After centrifugation (21,000 x g, 5 10 minutes), 0.6 ml of the hexane layer was transferred to a scintillation vial, scintillation fluid was added, and the radioactivity counted. Units of enzyme activity, and the lower limit of detection, were the same as defined above for IPP isomerase. In cases where high IPP isomerase activity interferes with measurement of FPP synthase activity, crude extract may be preincubated for 5 minutes in the presence of 5mM iodoacetamide to inhibit IPP 15 isomerase activity.

GGPP synthase assays. Crude extracts were prepared in 50 mM Tris-HCl buffer (pH 8.0) containing 2 mM dithiothreitol. GGPP synthase was assayed according to the procedure of Kuzuguchi et al. [J. Biol. Chem. 274:5888-5894 (1999)]. This assay is based on the same principle as described above for FPP synthase. Reaction mixtures (0.1 ml final volume) 20 contained 50 mM Tris-HCl buffer (pH 8.0), 2 mM dithiothreitol, 5 mM MgCl₂, 20 µM [1-¹⁴C]-IPP, 25 µM FPP and crude extract. All reaction conditions and subsequent treatment of samples for scintillation counting were identical to those described above for FPP synthase. Treatment of extract with iodoacetamide to inhibit IPP isomerase activity may also be used as above. Units of enzyme activity, and the lower limit of detection, were the same 25 as defined above for IPP isomerase.

Acetoacetyl-CoA reductase assays. Crude extracts are prepared in 50 mM Tris-HCl buffer (pH 7.5) containing 50 mM KCl and 5 mM dithiothreitol. Acetoacetyl-CoA reductase was assayed according to the procedure of Chohan and Copeland [Appl. Environ. Microbiol. 64:2859-2863 (1998)]. This assay measures the acetoacetyl-CoA-dependent oxidation of 30 NADPH spectrophotometrically at 340 nm. Reaction mixtures (1 ml) contain 50 mM Tris-HCl buffer (pH 8.5), 15 mM MgCl₂, 250 µM NADPH, and 100 µM acetoacetyl-CoA. Reactions are performed at in a quartz cuvette at 30°C and are initiated by the addition of acetoacetyl-CoA. Activity (expressed as U/mg protein) was calculated using a standard curve constructed with NADPH (1 unit of activity = 1 µmol NADPH oxidized/min). The 35 lower limit of detection of acetoacetyl-CoA reductase activity is about 0.01 U/mg.

Example 2: Taxonomic Reclassification of *Flavobacterium* sp. as *Paracoccus*

- This Example describes the taxonomic re-classification of the zeaxanthin-producing bacterium formerly designated *Flavobacterium* sp. strain R-1512 (ATCC 21588) as *Paracoccus* sp. strain R-1512 (ATCC 21588). A comprehensive genomic and
- 5 biochemical/physiological analysis was performed by the Belgian Coordinated Collections of Microorganisms/Laboratorium voor Microbiologie, Universiteit Gent (BCCMTM/LMG), using state-of-the-art methods currently accepted as the scientific standards for bacterial classification. Besides *Paracoccus* sp. strain R-1512, several other bacteria belonging to the genus *Paracoccus* were included in the study (summarized in Table 2).

10 **Table 2. Bacteria used in taxonomic study.**

Bacterium	Strain designation	Source or reference
<i>Paracoccus</i> sp.	R-1512 (ATCC 21588)	ATCC (environmental isolate) US 3,891,504
<i>Paracoccus</i> sp.	R1534	mutant derived from R-1512; US 6,087,152
<i>Paracoccus</i> sp.	R114	mutant derived from R-1512; This work
<i>Paracoccus</i> sp.	R-1506	environmental isolate; This work
<i>Paracoccus</i> sp.	MBIC3024	H. Kasai, Kamaishi Institute, Japan
<i>Paracoccus</i> sp.	MBIC3966	H. Kasai, Kamaishi Institute, Japan
<i>Paracoccus</i> sp.	MBIC4017	H. Kasai, Kamaishi Institute, Japan
<i>Paracoccus</i> sp.	MBIC4020	H. Kasai, Kamaishi Institute, Japan
<i>P. marcusii</i>	DSM 11574 ^T	Harker et al., infra
<i>P. carotinifaciens</i>	E-396 ^T	Tsubokura et al., infra
<i>P. solventivorans</i>	DSM 6637 ^T	Siller et. al., Int. J. Syst. Bacteriol. 46:1125-1130 (1996)

Strains R1534 and R114 are mutants derived from strain R-1512 by classical mutagenesis and screening for improved zeaxanthin production. The primary screening was accomplished by selecting the highest color intensity producing colonies. A secondary screening

- 48 -

was accomplished in liquid culture media by the HPLC methods according to Example 1. Strain R-1506 is an independent isolate obtained from the same initial screening of environmental microorganisms that provided strain R-1512. Strains MBIC3024, MBIC3966, MBIC4017 and MBIC4020 were identified as members of the genus *Paracoccus* by the 5 nucleotide sequences of their 16S rDNA genes (DNA sequences were deposited in the public EMBL database, see Table 5). *Paracoccus marcusii* DSM 11574^T and *Paracoccus carotinifaciens* E-396^T are recently described type strains of carotenoid-producing bacteria [Harker et al., Int. J. Syst. Bacteriol. 48:543-548 (1998); Tsubokura et al., Int. J. Syst. Bacteriol. 49:277-282 (1999)]. *Paracoccus solventivorans* DSM 6637^T was included as a 10 "control" strain, being a member of the genus *Paracoccus* but distantly related to the other bacteria used.

Preliminary experiments resulted in the following conclusions. Each of the methods set forth herein has a well-recognized ability to define taxonomic relatedness or relative degree of similarity between organisms. The methods and their use for delineating bacterial taxa 15 were described and compared in detail by Van Damme et al., Microbiological Reviews 60:407-438 (1996) and Janssen et al., Microbiology 142:1881-1893 (1996).

(1) Fatty acid analysis of the cell membranes of strains R1534 and R114 showed that the two strains were highly similar and indicated a taxonomic relatedness of these strains to *Paracoccus denitrificans* and *Rhodobacter capsulatus*.

20 (2) One-dimensional gel electrophoresis of cellular proteins showed a high similarity (i.e., a relatedness at the intra-species level) between R1534 and R114, but the profiles did not justify allocation of these strains to either *R. capsulatus* or *P. denitrificans*.

(3) DNA:DNA hybridization between strain R1534 and *R. capsulatus* LMG2962^T and *P. denitrificans* LMG4218^T confirmed that strain R1534 is neither *R. capsulatus* nor *P. denitrificans*.

25 (4) Sequencing of 16S rDNA genes from strains R1534 and R114 showed that these organisms belong to the genus *Paracoccus*, but that they represent a new species. The highest degree of sequence similarity was observed with the 16S rDNA gene of *Paracoccus* sp. strains MBIC3966, MBIC4020 and MBIC3024.

30 (5) DNA fingerprinting of strains R1534 and R-1512 using Amplified Fragment Length Polymorphism (AFLP™) showed high overall similarity of the genomic DNA from the two strains, indicating an infraspecific relatedness (i.e. AFLP™ can differentiate between two members of the same species).

- 49 -

In the following sections, the results and conclusions of the present comprehensive taxonomic study of *Paracoccus* sp. strain R-1512 (and its mutant derivatives R1534 and R114) are set forth.

16S rDNA sequencing and phylogenetic study. The bacteria set forth in Table 2 were grown in LMG medium 185 ((TSA) BBL 11768 supplemented where necessary with 1.5% Difco Bacto agar). Genomic DNA was prepared according to the protocol of Niemann et al. [J. Appl. Microbiol. 82:477-484 (1997)]. Genes coding for 16S rDNA were amplified from genomic DNA from strains R-1512, R1534, R114 and R-1506 by polymerase chain reaction (PCR) using the primers shown in Table 3.

10 **Table 3. Primers used for PCR amplification of DNA coding for 16S rDNA in *Paracoccus* sp. strains R-1512, R1534, R114, and R-1506.**

Primer name ^a	Sequence (5'→3')	SEQ ID NO	Position ^b
16F27	AGA GTT TGA TCC TGG CTC AG	SEQ ID NO:1	8-27
16F38	CTG GCT CAG GAC/T GAA CGC TG	SEQ ID NO:2	19-38
16R1522	AAG GAG GTG ATC CAG CCG CA	SEQ ID NO:3	1541-1522

^aF, forward primer; R, reverse primer. Forward primer 16F27 (Synonym: PA) was used for strains R1534 and R-1506, while forward primer 16F38 (Synonym: ARI C/T) was used for strains R-1512 and R114. The reverse primer 16R1522 (Synonym: PH) was used for all 15 strains.

^bHybridization position referring to *E. coli* 16S rDNA gene sequence numbering.

The PCR-amplified DNAs were purified using the Qiaquick PCR Purification Kit (Qiagen GmbH, Hilden, Germany). Complete sequencing was performed using an Applied Biosystems, Inc. 377 DNA Sequencer and the protocols of the manufacturer (Perkin-Elmer, 20 Applied Biosystems Division, Foster City, CA, USA) using the "ABI PRISM™ Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit (with AmpliTaq® DNA Polymerase, F_s)". The primers used for DNA sequencing are shown in Table 4.

Table 4. Primers used for sequencing PCR-amplified segments of genes coding for 16S rDNA in *Paracoccus* sp. strains R-1512, R1534, R114 and R-1506.

Primer name ^a /- Synonym	Sequence (5'→3')	SEQ ID NO	Position ^b

- 50 -

16F358/*Gamma	CTC CTA CGG GAG GCA GCA GT	SEQ ID NO:4	339-358
16F536/*PD	CAG CAG CCG CGG TAA TAC	SEQ ID NO:5	519-536
16F926/*O	AAC TCA AAG GAA TTG ACG G	SEQ ID NO:6	908-926
16F1112/*3	AGT CCC GCA ACG AGC GCA AC	SEQ ID NO:7	1093-1112
16F1241/*R	GCT ACA CAC GTG CTA CAA TG	SEQ ID NO:8	1222-1241
16R339/Gamma	ACT GCT GCC TCC CGT AGG AG	SEQ ID NO:9	358-339
16R519/PD	GTA TTA CCG CGG CTG CTG	SEQ ID NO:10	536-519
16R1093/3	GTT GCG CTC GTT GCG GGA CT	SEQ ID NO:11	1112-1093

^aF, forward primer; R, reverse primer.

^bHybridization position referring to *E. coli* 16S rDNA gene sequence numbering.

Five forward and three reverse primers were used to obtain a partial overlap of sequences, ensuring highly reliable assembled sequence data. Sequence assembly was performed
5 using the program AutoAssembler (Perkin-Elmer, Applied Biosystems Division, Foster City, CA, USA). Phylogenetic analysis was performed using the software package Gene-Compar™ (v. 2.0, Applied Maths B.V.B.A., Kortrijk, Belgium) after including the consensus sequences (from strains R-1512, R1534, R114 and R-1506) in an alignment of small ribosomal subunit sequences collected from the international nucleotide sequence library
10 EMBL. This alignment was pairwise calculated using an open gap penalty of 100% and a unit gap penalty of 0%. A similarity matrix was created by homology calculation with a gap penalty of 0% and after discarding unknown bases. A resulting tree was constructed using the neighbor-joining method.

The nucleotide sequence of the 16s rDNA gene from *Paracoccus* sp. strain R-1512 is
15 illustrated as SEQ ID NO:12. The distance matrix, presented as the percentage of 16S rDNA sequence similarity, between strain R-1512 and its closest relatives, is shown in Table 5. The sequences from strains R-1512 and its mutant derivatives R1534 and R114 were identical. The sequence from R-1506 differed by only one nucleotide from the sequence from latter strains. This demonstrated strains R-1512 and R-1506 are phylogenetically highly related and likely belong to the same species (confirmed by DNA:DNA hybridization, see below). Comparison of the R-1512 and R-1506 sequences with those publicly available at the EMBL library located R-1512 and R-1506 in the genus *Paracoccus*. However, the sequence similarities observed with all currently taxonomically validly described *Paracoccus* species was <97%, the value generally accepted as the limit for a
20 possible relatedness at the species level [Stackebrandt and Goebel, Int. J. Syst. Bacteriol.
25

44:846-849 (1994)]. This demonstrated that strains R-1512 (and its mutant derivatives) and R-1506 belong to one or two new *Paracoccus* species. Sequence similarities of >97% (significant for a possible relationship at the species level), were observed between four unnamed *Paracoccus* strains and strains R-1512, R1534, R114 and R-1506, suggesting that
 5 one or more of the unnamed (MBIC) strains may relate at the species level to strains R-1512 and R-1506. Based on cluster analysis (phylogenetic tree depicting the phylogenetic relatedness between *Paracoccus* sp. strains R-1512, R1534, R114, R-1506, MBIC3966, and other members of the genus *Paracoccus*), strains R-1512, R1534, R114, R-
 10 1506 and four unnamed *Paracoccus* strains (MBIC3024, MBIC3966, MBIC4017 and MBIC4020) were selected for DNA:DNA hybridization experiments to analyze species relatedness.

Table 5. Distance matrix, presented as the percentage of 16S rDNA sequence similarity, between *Paracoccus* sp. strain R-1512 and its closest relatives.

Strain ^a	EMBL Accession number	% Similarity
R-1512	-	100
R1534	-	100
R114	-	100
R-1506	-	99.9
<i>Paracoccus</i> sp. MBIC3966	AB018688	100
<i>Paracoccus</i> sp. MBIC3024	AB008115	98.2
<i>Paracoccus</i> sp. MBIC4020	AB025191	98.1
<i>Paracoccus</i> sp. MBIC4036	AB025192	97.0
<i>Paracoccus</i> sp. MBIC4017	AB025188	96.9
<i>Paracoccus</i> sp. MBIC4019	AB025190	96.8
<i>Paracoccus</i> sp. MBIC4018	AB025189	96.4
<i>Paracoccus marcusii</i> DSM 11574 ^T	Y12703	96.2
<i>Paracoccus carotinifaciens</i> E-396 ^T	AB006899	96.1
<i>Paracoccus solventivorans</i> DSM 6637 ^T	Y07705	95.4
<i>Paracoccus thiocyanaticus</i> THIO11 ^T	D32242	95.3
<i>Paracoccus aminophilus</i> JCM 7686 ^T	D32239	95.1
<i>Paracoccus alcaliphilus</i> JCM 7364 ^T	D32238	95.0

- 52 -

<i>Paracoccus pantotrophicus</i> ATCC 35512 ^T	Y16933	95.0
<i>Paracoccus denitrificans</i> ATCC 17741 ^T	Y16927	94.8
<i>Paracoccus versutus</i> IAM 12814 ^T	D32243	94.7
<i>Paracoccus kocurii</i> JCM 7684 ^T	D32241	94.6
<i>Paracoccus aminovorans</i> JCM 7685 ^T	D32240	94.4
<i>Paracoccus alkenifer</i> A901/1 ^T	Y13827	94.3
<i>Rhodobacter capsulatus</i> ATCC 11166 ^T	D16428	92.9

^aType strains are followed by a ^T

- DNA:DNA hybridization and determination of G+C content. The bacteria set forth in Table 5 were grown in LMG medium 185. Genomic DNA was prepared according to the protocol of Wilson [In Ausabel et al. (eds.), Current Protocols in Molecular Biology, Greene Publishing and Wiley Interscience, New York, 2.4.1-2.4.5 (1987)]. The G+C content of the DNA's was determined by HPLC according to Mesbach et al. [Int. J. Syst. Bacteriol. 39:159-167 (1989)] as modified by Logan et al. [Int. J. Syst. Evol. Microbiol. 50:1741-1753 (2000)]. Reported values are the mean of these measurements on the same DNA sample. DNA:DNA hybridizations were performed using the initial renaturation rate method as described by De Ley et al. [Eur. J. Biochem. 12:133-142 (1970)]. The hybridization temperature was 81.5°C. For this method, an average deviation of +/- 5.8% has been reported by Vauterin et al. [Int. J. Syst. Bacteriol. 45:472-489 (1995)]. The G+C content of the bacterial DNA's and the results of the DNA hybridization experiments are summarized in Table 6.
- 15 **Table 6. G+C content (mol %) of DNA from *Paracoccus* spp. strains and per cent DNA homology between the strains.**

Strain	%G+C	% DNA Homology							
R-1512	67.6	100							
R1534	67.7	96	100						
R114	67.5	100	97	100					
R-1506	67.5	94	90	88	100				
MBIC3024	65.4	31	nd ^a	nd	31	100			
MBIC3966	66.9	93	nd	nd	88	32	100		
MBIC4017	67.2	32	nd	nd	31	24	24	100	

MBIC4020	68.4	27	nd	nd	25	25	23	34	100
----------	------	----	----	----	----	----	----	----	-----

^anot determined

Strains R-1512, R1534, R114, R-1506 and MBIC3966 showed a DNA homology of >70% (the generally accepted limit for species delineation [Wayne et al., Int. J. Syst. Bacteriol. 37:463-464 (1987)], and therefore belong to the same species within the genus *Paracoccus*.

- 5 The G+C content of these five strains varied from 66.9%-67.7%, thus remaining within 1%, characteristic for a well defined species. On the other hand, the low DNA homology between strains MBIC3024, MBIC4017 and MBIC4020 and strains R-1512, R1534, R114, R-1506 and MBIC3966 showed that MBIC3024, MBIC4017 and MBIC4020 each belong to a different genomic species within the genus *Paracoccus*.

- 10 DNA fingerprinting using AFLP™. AFLP™ is a PCR-based technique for whole genome DNA fingerprinting via the selective amplification and selective visualization of restriction fragments [Vos et al., Nucleic Acids Research 23:4407-4414 (1995); Janssen et al., supra]. In this analysis, *Paracoccus* sp. strains R-1512, R1534, R114, R-1506, MBIC3966, and *Paracoccus marcusii* DSM 11574^T were compared to evaluate infraspecies relatedness. These
- 15 bacteria were grown in LMG medium 185. Genomic DNA from each of these bacteria was prepared according to the protocol of Wilson (supra). Purified DNA was digested by two restriction enzymes, a 4-base cutter and a 6-base cutter. In this way, a limited number of fragments with two different ends and of suitable size for efficient PCR were obtained. Adaptors (small double-stranded DNA molecules of 15-20 bp) containing one compatible
- 20 end were ligated to the appropriate "sticky" end of the restriction fragments. Both adaptors are restriction halfsite-specific, and have different sequences. These adaptors serve as binding sites for PCR primers. Here, the restriction enzymes used were *Apal* (a hexacutter, recognition sequence GGGCC/C) and *TaqI* (a tetracutter, recognition sequence T/GCA). The sequences of the adaptors ligated to the sticky ends generated by
- 25 cleavage with the restriction enzymes are shown in Table 7 (SEQ ID Nos:13-22). PCR was used for selective amplification of the restriction fragments. The PCR primers specifically annealed with the adaptor ends of the restriction fragments. Because the primers contain, at their 3' end, one so-called "selective base" that extends beyond the restriction site into the fragment, only those restriction fragments that have the appropriate complementary
- 30 sequence adjacent to the restriction site were amplified. The sequences of the six PCR primer combinations used are also shown in Table 7.

Table 7. Adaptors and PCR primers used for AFLP™ analysis.

- 54 -

	Sequence	SEQ ID NO
Adaptors corresponding to restriction enzyme <i>Apal</i>		
Adaptor 93A03	5'-TCGTAGACTGCGTACAGGCC-3'	SEQ ID NO:13
Adaptor 93A04	3'-CATCTGACGCATGT-5'	SEQ ID NO:14
Adaptors corresponding to restriction enzyme <i>TaqI</i>		
Adaptor 94A01	5'-GACGATGAGTCCTGAC-3'	SEQ ID NO:15
Adaptor 94A02	3'-TACTCAGGACTGGC-5'	SEQ ID NO:16
	Sequence	SEQ ID NO
Primer combination 1 (PC A)		
A01	5'GA <u>CTGCGTACAGGCCA</u> 3'	SEQ ID NO:17
T01	5'CGATGAGTCCTGAC <u>CGAA</u> 3'	SEQ ID NO:18
Primer combination 2 (PC B)		
A01	5'GA <u>CTGCGTACAGGCCA</u> 3'	SEQ ID NO:17
T02	5'CGATGAGTCCTGAC <u>CGAC</u> 3'	SEQ ID NO:19
Primer combination 3 (PC D)		
A02	5'GA <u>CTGCGTACAGGCC</u> 3'	SEQ ID NO:20
T01	5'CGATGAGTCCTGAC <u>CGAA</u> 3'	SEQ ID NO:18
Primer combination 4 (PC I)		
A03	5'GA <u>CTGCGTACAGGCCG</u> 3'	SEQ ID NO:21
T03	5'CGATGAGTCCTGAC <u>CGAG</u> 3'	SEQ ID NO:22
Primer combination 5 (PC G)		
A03	5'GA <u>CTGCGTACAGGCCG</u> 3'	SEQ ID NO:21
T01	5'CGATGAGTCCTGAC <u>CGAA</u> 3'	SEQ ID NO:18
Primer combination 6 (PC H)		
A03	5'GA <u>CTGCGTACAGGCCG</u> 3'	SEQ ID NO:21
T02	5'CGATGAGTCCTGAC <u>CGAC</u> 3'	SEQ ID NO:19

Following amplification, the PCR products were separated according to their length on a high resolution polyacrylamide gel using a DNA sequencer (ABI 377). Fragments that contained an adaptor specific for the restriction halfsite created by the 6-bp cutter were

- 55 -

visualized by autoradiography due to the 5'-end labeling of the corresponding primer with ^{32}P . The electrophoretic patterns were scanned and numerically analyzed with Gel-Compar™ 4.2 software (Applied Maths, B.V.B.A., Kortrijk, Belgium) and clustered using the Pearson curve matching coefficient and unweighted pair group averages linking 5 [clustering methods were reviewed by Sneath and Sokal, In: Numerical Taxonomy. Freeman & Son, San Francisco (1973)].

In all six primer combinations (PC A-H, Table 7), the DNA fingerprints of *Paracoccus* sp. strains R-1512, R1534 and R114 were highly similar if not identical. In cases where minor differences were observed, reproducibility was not evaluated. The high similarity or 10 identity among the three strains was expected as strains R1534 and R114 were derived from strain R-1512. With all primer combinations, strains R-1512, R1534 and R114 were clearly discriminated from strains R-1506 and MBIC3966, the latter two strains equally belonging to the new *Paracoccus* species. However, the fingerprints provide no clear indication that strains R-1512, R1534 and R114 are more related to either R-1506 or 15 MBIC3966. Under the conditions used, the five strains of the new species cluster at an average level of about 58% similarity (this value is the mean of the six values of the branching points of the new species in the six AFLP™ experiments (six primer combinations)), and the cluster can clearly be discriminated from the profile of *Paracoccus marcusii* DSM 11574^T, the type strain of a phylogenetically related carotenoid-producing 20 *Paracoccus* species. The mean similarity value of the six branching points for *Paracoccus marcusii* DSM 11574^T and the new species was about 11%.

Fatty acid analysis. The fatty acid composition of the cell membranes of *Paracoccus* sp. strains R-1512, R1534, R114, R-1506, MBIC3966 were compared to the type strains *P. marcusii* DSM 11574^T, *P. carotinifaciens* E-396^T and *P. solventivorans* DSM 6637^T. The 25 bacteria were grown for 24 hours at 28°C in LMG medium 185. The fatty acid compositions were determined by gas chromatography using the commercial system MIDI (Microbial Identification System, Inc., DE, USA). Extraction and analysis of fatty acids was performed according to the recommendations of the MIDI system. Table 8 summarizes the results for all strains tested. For the five strains of the new *Paracoccus* species (R-1512, 30 R1534, R114, R-1506, MBIC3966), the mean profile was calculated. All eight organisms showed a comparable fatty acid composition of their cell membranes, with 18:1 w7c as the major compound. Only minor differences in fatty acid composition were observed between the new *Paracoccus* species and the three type strains.

- 56 -

Utilization of carbon sources for growth. For testing the aerobic utilization of carbon sources, BIOLOG-SF-N Microplate microtiter plates (Biolog Inc., Hayward, CA, USA) containing 95 substrates were used with the exception that the substrate in well E6 was D,L-lactic acid methyl ester instead of the usual sodium salt of D,L-lactic acid. Cells from 5 each of the strains identified in Table 9 were grown for 24 hours at 28°C in LMG medium 12 (Marine Agar, Difco 0979). A cell suspension with a density equivalent to 0.5 McFarland units was prepared in sterile distilled water. From this suspension, 18 drops were transferred into 21 ml of AUX medium (API 20NE, bioMérieux, France) and mixed gently. 0.1 Milliliters of the suspension was transferred to each well of the BIOLOG 10 MicroPlates, and the plates were incubated at 30°C. Wells were visually checked for growth after 48 hours and after 6 days. Also, at 6 days the visual scoring was confirmed by reading the microtiter plates using the BIOLOG plate reader.

The results of the BIOLOG analysis are shown in Table 9. Growth (positive reaction) was determined as increased turbidity compared to the reference well without substrate. A 15 distinction was made between good growth (+), weak growth (\pm) and no growth (-). Results in parentheses are those obtained after 6 days if different from the results obtained after 48 hours. A question mark indicates an unclear result at 6 days. Of the 95 carbon sources tested, 12 could be used, and 47 could not be used, for growth by all five strains comprising the new *Paracoccus* species (R-1512, R1534, R114, R-1506 and MBIC3966). 20 These five strains gave variable growth responses to the remaining 36 substrates. The new *Paracoccus* species could be distinguished from the two other carotenoid-producing bacteria (*P. marcusii* DSM 11574^T and *P. carotinifaciens* E-396^T) by their inability to use seven carbon sources (adonitol, i-erythritol, gentiobiose, β -methylglucoside, D-sorbitol, xylitol and quinic acid). Two carbon sources that were utilized by all five members of the 25 new *Paracoccus* species (L-asparagine and L-aspartic acid) were not used for growth by *P. marcusii* DSM 11574^T.

Table 8. Fatty acid composition of cell membranes of *Paracoccus* sp. strains R-1512, R1534, R114, R-1506, MBIC3966 and three type strains of other *Paracoccus* species, i.e. *P. marcusii* DSM 11574^T, *P. carotinifaciens* E-396^T and *P. solventivorans* DSM 6637^T

	Mean % for:	% for:		
		DSM 11574 ^T	E-396 ^T	DSM 6637 ^T
Name	R-1512, R1534, R114, R-1506 and MBIC3966			
10:0 3OH	4.9 \pm 1.1	6.2	3.4	3.6

- 57 -

Unnamed 11.799	3.6 ± 0.5	4.9	2.8	3.0
Unnamed 15.275	1.5 ± 0.3	2.9	1.1	ND ^a
16:0	0.3 ± 0.2	ND	0.3	0.7
17:1 w8c	ND	ND	0.6	0.8
17:0	0.1 ± 0.1	ND	0.3	1.3
18:1 w7c	80.5 ± 1.8	80.3	84.0	79.0
	Mean % for:	% for:		
Name	R-1512, R1534, R114, R-1506 and MBIC3966	DSM 11574 ^T	E-396 ^T	DSM 6637 ^T
18:0	3.6 ± 0.4	2.6	5.2	6.6
18:0 3OH	0.6 ± 0.4	ND	ND	ND
19:0	ND	ND	ND	0.7
20:1 w7c	0.8 ± 0.2	ND	0.2	2.0
Summed feature 2	2.7 ± 0.4	3.0	2.1	2.6
Summed feature 3	0.7 ± 0.5	ND	0.2	ND
TOTAL	99.3	99.9	100.2	100.3

^a ND, not detected

Biochemical tests. Selected biochemical features were tested using the API 20NE strip (bioMérieux, France). Cells from each of the bacterial strains identified in Table 10 were grown for 24 hours at 28°C on LMG medium 12. Cell suspensions were prepared and strips inoculated according to the instructions of the manufacturer. Strips were incubated at 28°C and results determined after 24 and 48 hours. The results are summarized in Table 10. Of the nine features tested, only one (urease activity) gave a variable response among the five strains of the new *Paracoccus* species. These nine tests did not differentiate between the new *Paracoccus* species and *Paracoccus marcusii* DSM 11574^T and *P. carotinifaciens* E-396^T.

Table 9. Utilization of carbon sources for growth by *Paracoccus* spp. strains.

	R1512	R1534	R114	R1506	MBIC 3966	DSM 11574 ^T	E-396 ^T	DSM 6637 ^T

α -Cyclodextrin	-	-	-	-	-	-	-	-
Dextrin	-	-	-	-	-	-	-	- (\pm)
Glycogen	-	-	-	-	-	-	-	-
Tween 40	-	-	-	-	-	-	- $(?)$	-
Tween 80	-	-	-	-	-	-	-	-
GalNAc	-	-	-	-	-	-	-	-
	R1512	R1534	R114	R1506	MBIC 3966	DSM 11574 ^T	E-396 ^T	DSM 6637 ^T
GlucNAc	-	-	-	-	-	-	-	- $(?)$
Adonitol	-	-	-	-	-	+	+	-
L-Arabinose	-	-	-	-	-	+	-	+
D-Arabitol	+	+	+	+	$\pm(+)$	+	+	-
Cellobiose	$\pm(+)$	$\pm(+)$	- $(?)$	- $(+)$	- (\pm)	+	+	- $(+)$
i-Erythritol	-	-	-	-	-	+	+	-
D-Fructose	+	+	+	+	-	+	+	+
L-Fucose	-	-	-	-	-	-	+	-
D-Galactose	+	+	+	$\pm(+)$	$\pm(+)$	+	+	- (\pm)
Gentiobiose	-	-	-	-	-	+	+	- (\pm)
α -D-Glucose	+	+	+	$\pm(+)$	- $(+)$	+	$\pm(+)$	+
m-Inositol	+	+	+	- $(+)$	- $(+)$	+	- (\pm)	-
α -Lactose	+	$\pm(+)$	- $(+)$	- $(+)$	- $(+)$	+	+	$\pm(+)$
Lactulose	- (\pm)	- $(+)$	- $(+)$	- (\pm)	-	+	+	- $(+)$
Maltose	+	+	- $(+)$	- $(+)$	- (\pm)	+	+	- $(+)$
D-Mannitol	+	+	+	+	- $(+)$	+	+	- (\pm)
D-Mannose	+	+	+	+	- (\pm)	+	+	- $(+)$
D-Melibiose	+	+	+	- $(+)$	- $(+)$	+	+	- $(?)$
β -Methylgluc	-	-	-	-	-	+	+	+

- 59 -

D-Psicose	-(+)	\pm (+)	\pm (+)	-	-(+)	-	\pm	-
D-Raffinose	-	-	-	-	-	-(+)	+	-
L-Rhamnose	-	-	-	-	-	-	-	- (?)
D-Sorbitol	-	-	-	-	-	+	+	-
Sucrose	+	+	\pm (+)	-(+)	-	+	+	+
D-Trehalose	+	+	-(+)	-(+)	-(+)	+	+	+
	R1512	R1534	R114	R1506	MBIC 3966	DSM 11574 ^T	E-396 ^T	DSM 6637 ^T
Turanose	-(+)	-(+)	-	-	-	+	+	+
Xylitol	-	-	-	-	-	+	+	-
Methylpyruvate	\pm	-	\pm	- (?)	\pm	-	+	- (\pm)
MMsucc	\pm (+)	\pm	- (\pm)	-(+)	- (\pm)	-(+)	+	-
Acetic acid	-	-	\pm	-	-	-	-	+
Cis-aconitic acid	-	\pm	\pm	-	-	\pm	-	-
Citric acid	-	\pm	\pm	-	-	\pm	-	-
Formic acid	-	-	-	-	-	-	-	-
GalAlactone	- (\pm)	- (\pm)	- (\pm)	-	-	-	- (\pm)	- (?)
GalacturonicA	-	-	-	-	-	-(+)	- (\pm)	-
D-Gluconic acid	+	+	+	- (\pm)	- (\pm)	+	+	+
GlucosaminicA	-	-	-	-	-	-	-	-
GlucuronicA	\pm	+	+	- (\pm)	-	\pm (+)	-	-
AHBA	- (\pm)	-	- (\pm)	-	-(+)	-	-	-
BHBA	+	+	+	- (\pm)	\pm	-(+)	+	+
GHBA	-	-	-	-	-	-	-	-
PHPAA	-	-	-	-	-	-	-	- (+)
Itaconic acid	-	-	-	-	-	-	-	-

- 60 -

AKBA	-	-	-	-	-	-	-	-	-(±)
AKGA	-	-	-	-(±)	-(?)	-(±)	-(+)	-(±)	
AKVA	-	-	-	-	-	-	-	-	
LAME	-	-	-	-	-	-	-	-	
Malonic acid	-	-	-	-	-	-	-	-	
Propionic acid	-	±	±	-	-	±	+	+	
	R1512	R1534	R114	R1506	MBIC 3966	DSM 11574 ^T	E-396 ^T	DSM 6637 ^T	
Quinic acid	-	-	-	-	-	+	+	-	
SaccA	-(+)	±	-	-(±)	-	-	-	-	
Sebacic acid	-(+)	-(+)	-(+)	-(+)	-(±)	-	-(+)	-	
Succinic acid	-	-	-	-	-	-(+)	±	-(?)	
BromosuccA	-	-	-	-	-	±	-	-	
Succinamic acid	-	-	-	-	-	-(+)	-(+)	-	
Glucuronamide	-	-	-	-	-(±)	-	-	-	
Alaninamide	-	-	-	-	-	-(+)	+	-	
D-Alanine	-	-	-(+)	-	-	-	-	-	
L-Alanine	+	+	+	+	-	-(+)	+	+	
L-Alanyl-glycine	-(+)	-	-(+)	-	-	-	-(+)	-(?)	
L-Asparagine	+	+	±(+)	+	±(+)	-	+	-	
L-Aspartic acid	+	+	±(+)	-(+)	-(+)	-	+	-	
L-Glutamic acid	+	+	+	+	±(+)	-(+)	+	-(+)	
GAA	-	-	-	-(±)	-	-	-	-	
GGA	-(?)	-	-	-(?)	-	-	-(±)	-	
L-Histidine	-	-	-	-	-	-(?)	-	+	
HydPro	-	-	-	-	-	-	-	-	+
L-Leucine	-(±)	-(+)	-(+)	-(+)	-	-(+)	-(?)	-(+)	

L-Ornithine	-	-(+)	\pm (+)	-(\pm)	-	-	-(+)	-
L-Phenylalanine	-	-	-	-	-	-	-	-
L-Proline	+	+	+	+	-	-(+)	+	+
PyroGluA	+	+	+	+	\pm (+)	-(+)	+	-
D-Serine	-	-	-	-	-	-	-	-
L-Serine	\pm	\pm (+)	-(+)	-(\pm)	-(+)	-(+)	-(+)	+
	R1512	R1534	R114	R1506	MBIC 3966	DSM 11574 ^T	E-396 ^T	DSM 6637 ^T
L-Threonine	-	-	-	-	-	-(+)	-	-
D,L-Carnitine	-	-	-	-	-	-	-	-
GABA	-	-	-	-	-(+)	-	-(+)	-(+)
Urocanic acid	-	-	-	-	-	-	-	-(+)
Inosine	-	-(\pm)	-	-	-	\pm (+)	-(+)	-(+)
Uridine	-	-	-	-	-	\pm (+)	-(+)	-
Thymidine	-	-	-	-	-	\pm (+)	-(+)	-
PEA	-	-	-	-	-	-	-	-
Putrescine	-	-	-	-	-	-	-	-
2-Aminoethanol	-	-	-	-	-	-	-	-
2,3-Butanediol	-	-	-	-	-	-	-	-
Glycerol	+	+	+	-(+)	-	+	+	-
GlycP	-	-	-	-	\pm (+)	-	-	-
Gluc-1-P	-	-	-	-	-	-	\pm (+)	-
Gluc-6-P	-	-	-	-	-	-	-	-

GalNAc: N-Acetyl-D-galactosamine; GlucNAc: N-Acetyl-D-glucosamine; β -Methylgluc:

β -Methylglucoside; MMSucc: Mono-methylsuccinate; GalAlactone: D-Galactonic acid

lactone; GalacturonicA: D-Galacturonic acid; GlucosaminicA: D-Glucosaminic acid;

GlucuronicA: D-Glucuronic acid; AHBA: α -Hydroxybutyric acid; BHBA: β -Hydroxy-

butyric acid; GHBA: γ -Hydroxybutyric acid; PHPAA: p-Hydroxyphenylacetic acid; AKBA:

- 62 -

- α -Ketobutyric acid; AKGA: α -Ketoglutaric acid; AKVA: α -Ketovaleric acid; LAME: D,L-Lactic acid methyl ester; SaccA: D-Saccharic acid; BromosuccA: Bromosuccinic acid; GAA: Glycyl-L-aspartic acid; GGA: Glycyl-L-glutamic acid; HydPro: Hydroxy-L-proline; PyroGluA: L-Pyroglutamic acid; GABA: γ -Aminobutyric acid; PEA: Phenylethylamine;
- 5 GlycP: D,L- α -Glycerolphosphate; Gluc-1-P: Glucose-1-phosphate; Gluc-6-P: Glucose-6-phosphate

Table 10. Biochemical features of *Paracoccus* spp. strains: 12 = R1512; 34 = R1534; 14 = R114, 06 = R1506; 66 = MBIC3966; 74 = DSM 11574^T, 96 = E-396^T, 37 = DSM 6637^T

	12	34	14	06	66	74	96	37
Reduction nitrate to nitrite	-	-	-	-	-	-	-	+
Reduction nitrate to nitrogen	-	-	-	-	-	-	-	+
Indole from tryptophan	-	-	-	-	-	-	-	-
Fermentation of glucose	-	-	-	-	-	-	-	-
Arginine hydrolase	-	-	-	-	-	-	-	-
Urease	S / + 5	-	-	S / + 5	+	-	-	-
Esculin hydrolysis ^a	weak	S / + 5	S / + 5	+	+	+	+	+
Gelatine hydrolysis ^b	-	-	-	-	-	-	-	-
β -Galactosidase	+	+	+	+	+	+	+	-

^a: β -glucosidase; ^b: protease; S / + 5: Slow + 5 days

- 10 Physiological tests. Several physiological and morphological tests were performed on the five strains of the new *Paracoccus* species, along with *Paracoccus marcusii* DSM 11574^T, *Paracoccus carotinifaciens* E-396^T and *Paracoccus solventivorans* DSM 6637^T. The methods used for each test were as follows.

- Temperature range for growth.* Cells were grown for 24 hours at 28°C on LMG medium 12.
- 15 A cell suspension with a density of between 1-2 McFarland units was prepared in sterile distilled water. From this suspension, 3 drops were transferred onto the agar surface of LMG medium 12. One drop was diluted by streaking, the other 2 drops were left undisturbed. The plates were incubated under aerobic conditions at 10°C, 25°C, 30°C, 33°C, 37°C and 40°C, and checked for growth after 24 hours, 48 hours and 5 days. Growth was
- 20 determined as visual growth (confluent in the drops and as colonies in the streaks with

- 63 -

diluted inoculum) compared to the growth at 30°C (i.e., the "control"). Scoring was done (vs. the control plate) as follows; better growth (++) , good (equivalent to the control) growth (+), weaker growth (±), poor growth (±), and no growth (-). Results in parentheses are those observed in the streaks if different from the confluent growth in the undisturbed drops.

Salt tolerance. Cells were grown for 24 hours at 28°C on LMG medium 12. A cell suspension with a density of between 1-2 McFarland units was prepared in sterile distilled water. From this suspension, 3 drops were transferred onto the agar surface of LMG medium 12 supplemented with NaCl to reach final concentrations of 3%, 6% and 8%. One drop was diluted by streaking, the other 2 drops were left undisturbed. The plates were incubated under aerobic conditions at 28°C and checked for growth after 24 hours, 48 hours and 5 days. Growth was determined as visual growth (confluent in the drops and as colonies in the streaks with diluted inoculum) compared to the growth without added NaCl (control). Scoring was done (vs. the control plate) as follows; better growth (++) , good (equivalent to the control) growth (+), weaker growth (±), poor growth (±), and no growth (-). Results in parentheses are those observed in the streaks if different from the confluent growth in the undisturbed drops.

pH Range for growth. Cells were grown for 24 hours at 28°C in LMG medium 12. A cell suspension with a density of between 1-2 McFarland units was prepared in sterile distilled water. From this suspension, 3 drops were transferred into tubes containing 10 ml liquid LMG medium 12 with modified pH, giving final pH values after autoclaving of pH 6.1, pH 6.3, pH 7.0, pH 7.7, pH 8.1 and pH 9.1. The liquid cultures were incubated aerobically (with shaking) at 28°C. Growth was checked at 24 hours, 48 hours, 3 days and 6 days. Growth was determined as increased turbidity (measured as % transmission using the BIOLOG turbidimeter) compared to growth at pH 7.0 (control). Scoring was done (vs. the control) as follows; better growth (++) , good (equivalent to the control) growth (+), weaker growth (±), poor growth (±), and no growth (-).

Starch hydrolysis. Cells were grown for 24 hours at 28°C on LMG medium 12 plates. A loopful of cells was taken from the plate and transferred as one streak onto the agar surface of LMG medium 12 supplemented with 0.2% soluble starch. Plates were then incubated under aerobic conditions at 28°C. When the strains had developed good growth (after 48 hours), the plate was flooded with lugol solution (0.5% I₂ and 1% KI in distilled water). Hydrolysis was determined as a clear zone alongside the growth (in contrast to the blue color of the agar where starch was not hydrolyzed).

- 64 -

Denitrification. Cells were grown for 24 hours at 28°C on LMG medium 12 plates. A loopful of cells was taken from the plate and stabbed into tubes containing semi-solid (0.1% agar) LMG medium 12 supplemented with 1% KNO₃. The plates were incubated at 28°C for 5 days. Denitrification (N₂ from nitrate) was determined as gas formation
5 alongside the stab.

Growth under anaerobiosis without electron acceptor added. Cells were grown for 24 hours at 28°C on LMG medium 12 plates. A loopful of cells was taken from the plate and streaked onto the agar surface of LMG medium 12. The agar plates were incubated under anaerobic conditions (ca. 10% CO₂ + ca. 90% N₂) at 30°C. Plates were checked for growth
10 after 24 hours and after 5 days. Growth was determined visually and compared to the aerobic (control) condition. Scoring was done (vs. the control) as follows; better growth (++)
and no growth (-).
C

Growth under anaerobic conditions with glucose added (fermentation). Cells were grown for
15 24 hours at 28°C on LMG medium 12 plates. A loopful of cells was taken from the plate and stabbed into tubes containing the basal agar medium of Hugh and Leifson [J.
Bacteriol. 66:24-26 (1953)]. Paraffin oil was added to the top of the medium, and the tubes were incubated at 30°C. Tubes were checked for growth and acid formation after 48 hours and after 5 days. Growth was determined visually. Scoring was done as follows;
20 good growth (+), poor growth (±), and no growth (-).

Growth under anaerobic conditions with KNO₃ as electron acceptor. Cells were grown for 24 hours at 28°C on LMG medium 12 plates. A loopful of cells was taken from the plate and streaked onto the agar surface of LMG medium 12 supplemented with 0.1% KNO₃. The plates were incubated under anaerobic conditions (ca. 10% CO₂ + ca. 90% N₂) at 30°C,
25 and checked for growth after 3 days. Growth was determined as visual growth compared to the aerobic (control) condition. Scoring was done (vs. the control) as follows; better growth (++)
and no growth (-).
C

Catalase and oxidase reactions. Cells were grown for 24 hours at 28°C on LMG medium 12
30 plates. A positive result for catalase activity was the production of gas bubbles after suspending a colony in one drop of 10% H₂O₂. A positive result for oxidase activity was the development of a purple-red color after rubbing a colony on filter paper soaked with 1% tetramethylparaphenylene.

- 65 -

Colony pigmentation. Cells were grown for 5 days at 28°C on LMG medium 12. Color of colonies was observed visually.

Cell morphology and motility. Cells were grown for 24 hours at 28°C on LMG medium 12. Cell suspensions were made in sterile saline. Cell morphology and motility were observed 5 microscopically using an Olympus light microscope equipped with phase contrast optics (magnification 1000x).

The results of the physiological and morphological tests are summarized in Table 11. The five strains of the new *Paracoccus* species responded essentially identically in all physiological and morphological tests performed. The tests that gave identical responses for all five 10 strains of the new *Paracoccus* species and that allowed discrimination of these organisms from *Paracoccus marcusii* DSM 11574^T and/or *Paracoccus carotinifaciens* E-396^T were: growth at 40°C, growth with 8% NaCl, growth at pH 9.1, and colony pigmentation.

Zeaxanthin production in strains R-1512, R1534, R114 and R-1506 strains. Strains R-1512, R1534, R114, and R-1506 were grown in ME medium, which contains (per liter 15 distilled water): 5 g glucose, 10 g yeast extract, 10 g tryptone, 30 g NaCl and 5 g MgSO₄·7H₂O. The pH of the medium was adjusted to 7.2 with 5N NaOH before sterilizing by autoclaving. All cultures (25-ml volume in 250-ml baffled Erlenmeyer flasks with plastic caps) were grown at 28°C with shaking at 200 rpm. Seed cultures were inoculated from frozen glycerolized stocks and grown overnight. Aliquots were 20 transferred to the experimental flasks to achieve an initial optical density at 660 nm (OD₆₆₀) of 0.16. Cultures were then grown at 28°C with shaking at 200 rpm. Growth was monitored throughout the cultivation and at 6, 10 (or 15 for strain R114), and 24 hours, (an aliquot of the culture was removed for analysis of carotenoids by the method described in Example 1.

25 The doubling times of strains R-1512, R1534 and R-1506 under these conditions were 0.85 hours, 1.15 hours and 1.05 hours, respectively. Strain R114 reproducibly exhibited a bi-phasic growth profile; the doubling time of strain R114 in the initial phase was 1.4 hours while the doubling time in the second phase was 3.2 hours.

Table 12 shows the zeaxanthin production and Specific Formation (zeaxanthin production 30 normalized to OD₆₆₀) by the *Paracoccus* sp. strains in ME medium. The data are averages of four independent experiments, and within each experiment each strain was tested in duplicate flasks. The improved zeaxanthin production in the classically-derived mutant strains R1534 and R114 compared to the parental strain R-1512 is clearly shown. Zea-

- 66 -

xanthin production by strain R-1506 was approximately the same as strain R-1512. No other carotenoids were detected in any of the cultures.

(

C

- 67 -

Table 11. Physiological characteristics of *Paracoccus* spp. strains: 12 = R1512; 34 = R1534; 14 = R114, 06 = R1506; 66 = MBIC3966; 74 = DSM 11574^T, 96 = E-396^T, 37 = DSM 6637^T

Time [days]	12	34	14	06	66	74	96	37
Growth at 10°C								
1	-	-	-	-	-	-	-	-
5	± (±)	± (±)	± (-)	± (±)	± (-)	± (±)	± (±)	± (±)
Growth at 25°C								
1	+	+	+	+	+ (±)	+ (±)	+ (±)	+ (-)
5	+	+	+	+	+	+	+	+
Growth at 30°C								
1	+	+	+	+	+	+	+	+
5	+	+	+	+	+	+	+	+
Growth at 33°C								
1	+	+	+	+	+	+	+	+
5	+	+	+	+	+	+	+	+
Growth at 37°C								
1	+	+	+ (±)	+	+	± (-)	± (-)	+
5	+	+	+	+	+	± (-)	± (±)	+
Growth at 40°C								
1	+	+ (±)	+ (-)	+ (±)	± (-)	-	-	+ (*)
5	+	+ (±)	+ (-)	+	+ (-)	-	-	+ (*)
Growth with 3% NaCl								
1	+	+	+	+	+	+	+	±
5	+	+	+	+	+	+	+	+
Growth with 6% NaCl								
1	+ (±)	± (±)	± (±)	+	± (±)	± (-)	± (-)	-
5	+	+	+	+	+ (*)	+ (±)	+ (±)	-

- 68 -

Time [days]	12	34	14	06	66	74	96	37
Growth with 8% NaCl								
1	+ (±)	± (±)	± (-)	+ (±)	± (±)	-	-	-
5	+	+	+	+	+ (*)	± (-)	± (-)	-
Growth at pH 6.1								
1	+	+	+	+	-	-	-	-
6	+	+	+	+	+	+	+	+
Growth at pH 6.3								
1	+	+	+	+	+	±	+	±
6	+	+	+	+	+	+	+	+
Growth at pH 7.0								
1	+	+	+	+	+	+	+	+
6	+	+	+	+	+	+	+	+
Growth at pH 7.7								
1	+	+	+	+	+	±	±	±
6	+	+	+	+	+	+	+	+
Growth at pH 8.1								
1	+	+	+	+	+	-	-	±
6	+	+	+	+	+	+	+	+
Growth at pH 9.1								
1	±	+	-	-	+	-	-	-
6	+	+	+	+	+	-	+	+
Starch hydrolysis								
	-	-	-	-	-	-	-	-
Denitrification								
	-	-	-	-	-	-	-	+
Growth in anaerobiosis without electron acceptor added								
	-	-	-	-	-	-	-	-

- 69 -

	12	34	14	06	66	74	96	37
Growth in anaerobiosis with glucose added (fermentation)								
	-	-	-	-	-	-	-	±
Growth in anaerobiosis with KNO ₃ added								
	-	-	-	-	-	-	-	-
Catalase reaction								
	+	+	+	+	+	+	+	+
Oxidase reaction								
	+	+	+	+	+	+	+	+
Gram stain								
	-	-	-	-	-	-	-	-
Motility								
	-	-	-	-	-	-	-	-
Colony pigmentation								
	Y - O	Y - O	Y - O	Y - O	Y - O	O - P	O - P	P Y
Cell morphology								
	S to C	S to C	S to C	C	S to C	S	S	S
Cell dimensions (μm)								
	0.8 x 1.2	0.8 x 1.2	0.8 x 1.2	0.9 x 1.1	0.8 x 1.2 to 1.5	0.8 x 1.5 to 2.0	0.9 x 2.0 to 2.5	0.8 x 1.5 to 2.0

Y - O: yellow-orange; O - P: orange-pink; P Y: pale yellow; S to C: short rod to coccoid; S: short rod; C: coccoid

Table 12. Zeaxanthin production by *Paracoccus* sp. strains R-1512, R1534, R114 and R-1506.

Strain	Time (h)	Zeaxanthin (mg/l)		Specific Formation (mg zeaxanthin/OD ₆₆₀)	
		Average	Standard Deviation	Average	Standard Deviation
R-1512	6	0.23	0.10	0.10	0.04
	10	2.05	0.70	0.25	0.08
	24	3.78	0.59	0.38	0.06
R1534	6	0.75	0.10	0.26	0.02
	10	3.45	0.57	0.43	0.07
	24	9.13	0.97	0.95	0.06
R114	6	0.65	0.17	0.86	0.24
	15	7.53	1.12	1.13	0.21
	24	19.7	1.82	2.68	0.20
R-1506	6	0.13	0.06	0.07	0.01
	10	1.35	0.31	0.19	0.04
	24	3.55	0.68	0.38	0.07

Example 3: IPP Biosynthesis via the Mevalonate Pathway in the Zeaxanthin-Producing *Paracoccus* sp. strain R114.

In order to determine the biosynthetic origin (*i.e.*, the mevalonate or DXP pathway) of isoprenoid precursors in *Paracoccus* sp. strain R114, a “retrobiosynthesis” approach [Eisenreich and Bacher, In: Setlow (ed.), Genetic Engineering, Principles and Methods, Kluwer Academic/Plenum Publishers, New York 22:121-153 (2000)] was taken. This predictive approach for data analysis permits the unequivocal assessment of glucose catabolism from the analysis of a single down-stream natural product. In the present work, this involved growth of the bacterium in media containing various binary mixtures of unlabeled glucose and specific ¹³C-labeled glucoses, followed by purification of the zeaxanthin produced and analysis of the labeling patterns by NMR spectroscopy. Details of the methods used and the experimental results are given below.

Growth of *Paracoccus* sp. strain R114 for ¹³C labeling experiments. Unlabelled D-glucose monohydrate was purchased from Fluka (Milwaukee, WI, USA). [U-¹³C₆]-D-Glucose was purchased from Isotec (Miamisburg, OH, USA), while [1-¹³C₁] D-glucose, [2-¹³C₁] D-glucose and [6-¹³C₁] D-glucose were from Cambridge Isotope Laboratories (Andover, MA, USA). Yeast extract and peptone (from casein, pancreatically digested) were purchased

- 71 -

from EM Science (Gibbstown, NJ, USA). All other salts and solvents were analytical grade and were purchased from standard chemicals suppliers.

All cultures were initiated from frozen cell suspensions (cell density of 12 OD₆₆₀ units, 25% glycerol, stored at -70°C). One ml of thawed cell suspension was used to inoculate pre-
5 cultures (500-ml baffled shake flasks) containing 100 ml of 362F/2 medium having the following composition: 30 g/l D-glucose, 10 g/l yeast extract, 10 g/l peptone, 5 g/l NaCl, 2.5 g/l MgSO₄·7H₂O, 0.75 g/l (NH₄)₂HPO₄, 0.625 g/l K₂HPO₄, 187.5 mg/l CaCl₂·2H₂O, 0.2 g/l (NH₄)₂Fe(SO₄)₂·6H₂O, 15 mg/l ZnSO₄·7H₂O, 12.5 mg/l FeCl₃·6H₂O, 5 mg/l MnSO₄·H₂O, 0.5 mg/l NiSO₄·6H₂O, 15 mg/l Na-EDTA and 9.375 µl/l HCl (37% stock
10 solution). The initial pH of the medium was 7.2.

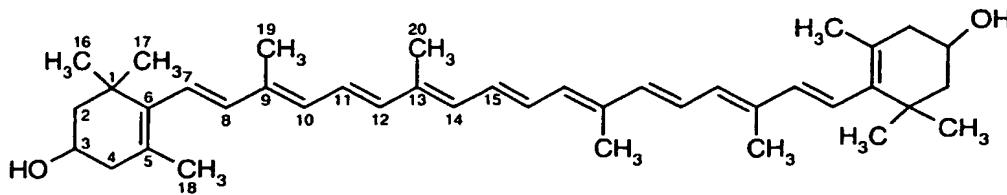
The pre-culture was incubated at 28°C with shaking at 200 rpm for 24 h, after which time the OD₆₆₀ was about 22 absorbance units. The main cultures were grown in Bioflo 3000 bioreactors (New Brunswick Scientific, Edison, NJ, USA) containing 362F/2 medium containing the following composition: 30 g/l total D-glucose (see below for ratios of ¹³C-labeled:unlabeled glucose), 20 g/l yeast extract, 10 g/l peptone, 10 g/l NaCl, 5 g/l MgSO₄·7H₂O, 1.5 g/l (NH₄)₂HPO₄, 1.25 g/l K₂HPO₄, 0.4 g/l (NH₄)₂Fe(SO₄)₂·6H₂O, 375 mg/l CaCl₂·2H₂O, 30 mg/l ZnSO₄·7H₂O, 25 mg/l FeCl₃·6H₂O, 10 mg/l MnSO₄·H₂O, 1 mg/l NiSO₄·6H₂O, 30 mg/l Na-EDTA and 18.75 µl/l HCl (37% stock solution). The amounts of each ¹³C-labeled glucose used (expressed as a percentage of the total 30 g/l glucose in the medium) in four separate experiments were: Condition 1, 4% [U-¹³C₆] D-glucose; Condition 2, 50% [¹³C₁] D-glucose; Condition 3, 25% [2-¹³C₁] D-glucose + 1% [U-¹³C₆] D-glucose; Condition 4, 25% [6-¹³C₁] D-glucose + 1% [U-¹³C₆] D-glucose. A control with only unlabeled glucose was also included. For Conditions 1 and 2 (and the unlabeled control), the culture volume was 2 l, while the culture volume for Conditions 3 and 4 was 1
20 l. The bioreactors were inoculated with pre-culture (20 ml/l initial volume) and cultivation proceeded for 22-24 hours, at which time no glucose was left in the medium. Cultivation conditions were: 28°C, pH 7.2 (controlled with 25% H₃PO₄ and 28% NH₄OH), dissolved oxygen controlled (in a cascade with agitation) at a minimum of 40%, agitation rate
25 and aeration rate 300 rpm (minimum) and 1 vvm, respectively.

30 Purification of zeaxanthin. At the end of the cultivations, the cultures were cooled down to 15°C. Five hundred ml of absolute ethanol was added per liter of culture and stirring was continued at 100 rpm for 20 min. The treated culture was centrifuged for 20 min. at 5000 x g, and the supernatant was discarded. The wet pellet was then extracted with 5 volumes of THF for 20 min. with stirring. The extracted mixture was centrifuged, the

- 72 -

supernatant saved and the resulting pellet extracted a second time with 1 volume THF under the same conditions and again centrifuged. The supernatants (extracts) were combined and concentrated to 50 ml by rotary evaporation. Five milliliters of hexane were added to the concentrated THF solution. After mixing, the system formed an emulsion 5 that could be separated by centrifugation. The aqueous phase was collected, diluted with an equal volume of saturated NaCl solution and re-extracted with dichloromethane. The dichloromethane phase was collected and combined with the THF/hexane phase. The mixture of organic extracts was concentrated again in a rotary evaporator to remove dichloromethane. The solution was then applied to a silica gel column and eluted with a 10 mixture of n-hexane and ether (1:1). A small light yellow band eluted first and was discarded. The main zeaxanthin product eluted in a broad band that moved slowly in the column. About 2 liters of solvent was needed to elute the main band completely. The eluate was collected in a round-bottomed flask and the solvent was removed by rotary evaporation at 40°C. The residue was dissolved in a small amount of dichloroethane at 40°C 15 and the solution was then allowed to cool slowly. Hexane was added to the mixture drop-wise until a turbidity was observed. The crystallization was complete within 48 hours at 4°C. The crystals were collected on a paper filter, washed with cold methanol and dried under vacuum.

NMR studies. Zeaxanthin was analyzed by NMR spectroscopy. For reference, the 20 chemical structure of zeaxanthin is illustrated in the following formula



¹H-NMR and ¹³C-NMR spectra were recorded at 500.13 MHz and 125.6 MHz, respectively, with a Bruker DRX 500 spectrometer. Acquisition and processing parameters for one-dimensional experiments and two-dimensional INADEQUATE experiments were according to standard Bruker software (XWINNMR). The solvent was deuterated chloroform. 25 The chemical shifts were referenced to solvent signals.

¹³C NMR spectra of the isotope labeled zeaxanthin samples and of the zeaxanthin sample at natural ¹³C abundance were recorded under the same experimental conditions. Integrals were determined for every ¹³C NMR signal, and the signal integral for each 30 respective carbon atom in the labeled compound was referenced to that of the natural abundance material, thus affording relative ¹³C abundances for each position in the labeled

- 73 -

molecular species. The relative abundances were then converted into absolute abundances from ^{13}C coupling satellites in the ^1H NMR signal of H-18 at 1.71 ppm. In the ^{13}C NMR spectrum of the multiply-labeled zeaxanthin sample each satellite was integrated separately. The integral of each respective satellite pair was then referenced to the total 5 signal integral of a given carbon atom. Zeaxanthin comprises a total of eight isoprenoid moieties (2 DMAPP units and 6 IPP units); only 20 ^{13}C NMR signals are observed due to chemical shift degeneracy.

In the experiment with the mixture of [$\text{U}-^{13}\text{C}_6$] glucose and unlabeled glucose (1:7.5; w/w), all carbon atoms of zeaxanthin were labeled and showed satellites due to $^{13}\text{C}^{13}\text{C}$ couplings 10 (Table 13). The signals of 4 carbon atoms have intense satellites due to $^{13}\text{C}^{13}\text{C}$ couplings (61.2 ± 0.6 % in the global NMR signal intensity of a given atom, Table 13). The signal accounting for the methyl atoms C-17/C-17' displayed only weak ^{13}C -coupled satellites at a relative intensity of 6%. The central signals represent material derived from unlabeled 15 glucose. The signals showed no evidence of long-range coupling. Carbon connectivity was easily gleaned from $^{13}\text{C}^{13}\text{C}$ coupling constants (Table 13) and from two-dimensional INADEQUATE experiments.

Three of the carbon atoms acquired label from [6- $^{13}\text{C}_1$] glucose. The other two carbons were labeled from [2- $^{13}\text{C}_1$] glucose. No significant amounts of label were contributed to zeaxanthin by [1- $^{13}\text{C}_1$] glucose.

20 The ^{13}C abundance for all non-isochronous carbon atoms was determined by comparison with spectra of unlabeled zeaxanthin and by evaluation of the $^1\text{H}^{13}\text{C}$ coupling satellites in ^1H NMR spectra (Table 13). The fraction of jointly transferred carbon atom pairs in the experiment with [$\text{U}-^{13}\text{C}_6$] glucose was determined by integration of the coupling satellites.

The labeling patterns of the IPP building block can be reconstructed accurately as shown 25 by the standard deviations found for the reconstructed IPP precursor. The reconstructed labeling patterns of DMAPP and IPP were identical within the experimental limits.

Table 13. NMR results for ^{13}C labeled zeaxanthin produced by *Paracoccus* sp. strain R114 supplied with ^{13}C labeled glucoses.

Position	δ - ^{13}C ppm	J_{CC} , Hz	^{13}C -labeled glucose precursor				
			[1- ^{13}C]-	[2- ^{13}C]-	[6- ^{13}C]-	% ^{13}C	% ^{13}C
	134.08	44.2 (18, 18')	% ^{13}C	% ^{13}C	% ^{13}C	% ^{13}C	% ^{13}C
1, 1'	37.13	36.0 (16,16')	1.10	10.71	2.22	3.47	61.2
2, 2'	48.46	35.8 (3,3')	1.20	2.58	10.27	3.65	61.1
3, 3'	65.10	35.8 (2,2')	1.12	12.47	2.38	3.64	60.4
4, 4'	42.57	37.1	1.27	2.59	10.63	3.89	8.4
5, 5'	126.17	44.2 (18, 18')	1.14	12.45	3.19	3.68	61.1
6, 6'	137.77	56.4 (7,7')	1.30	2.15	9.98	3.60	60.4
7, 7'	125.59	56.2 (6,6')	1.12	10.11	2.82	4.09	61.4
8, 8'	138.50	71.6, 55.7	1.28	2.24	9.95	3.92	4.3, 5.0
9, 9'	135.69	43.1 (19,19')	1.12	9.53	2.95	3.84	61.7
10, 10'	131.31	59.7 (11,11')	1.21	3.18	9.61	3.80	61.1
11, 11'	124.93	59.7 (10,10')	1.10	8.79	2.70	4.02	61.0
12, 12'	137.57	70.5	1.20	2.01	8.80	3.59	5.1
13, 13'	136.48	43.1 (20,20')	1.12	9.86	3.59	3.93	61.4
14, 14'	132.60	60.4 (15,15')	1.21	2.83	10.51	3.77	59.5
15, 15'	130.08	60.4 (14,14')	1.12	9.18	3.33	4.02	61.2
16, 16'	30.26	36.3 (1,1')	1.27	3.19	12.31	3.91	62.0
17, 17'	28.73	34.9 (1,1')	1.30	3.43	12.31	3.88	6.0
18, 18'	21.62	44.2 (5,5')	1.27	3.01	11.66	3.70	62.0
19, 19'	12.82	43.1 (9,9')	1.29	3.12	11.64	3.86	62.3
20, 20'	12.75	42.9 (13,13')	1.33	3.21	11.99	3.75	62.1

The experimental labeling patterns determined above can be compared with various predictions, taking into account not only the mevalonate pathway vs. the DXP pathway for isoprenoid biosynthesis, but also different pathways of glucose metabolism. Eubacteria typically utilize glucose primarily via the glycolytic pathway or via the Entner-Doudoroff pathway. Glycolysis generates two triose phosphate molecules from glucose. The C-1 and C-6 of glucose are both diverted to the 3-position of the triose phosphates produced during glycolysis. On the other hand, in the Entner-Doudoroff pathway, glucose is converted to a mixture of glyceraldehyde 3-phosphate and pyruvate. The C-1 of glucose is

- 75 -

exclusively diverted to C-1 of pyruvate, and the C-6 of glucose is exclusively diverted to C-3 of glyceraldehyde 3-phosphate.

Intermediates and products of the glycolytic and Entner-Doudoroff pathways serve as starting material for both isoprenoid biosynthetic pathways. With regard to the mevalon-

5 ate pathway, pyruvate as well as triose phosphate can be converted to the precursor acetyl-CoA. Glucose catabolism via the glycolytic pathway diverts label from C-1 as well as C-6 of glucose to the methyl group of acetyl-CoA. Glucose catabolism via the Entner-Doudoroff pathway results in loss of C-1 from glucose during the transformation of pyruvate to acetyl-CoA.

10 The experimentally observed enrichment and ^{13}C - ^{13}C coupling patterns of the zeaxanthin produced by *Paracoccus* sp. strain R114 were in perfect agreement with the labeling pattern required for zeaxanthin biosynthesis by the combination of the Entner-Doudoroff pathway and the mevalonate pathway. If both the glycolytic and Entner-Doudoroff pathways had been simultaneously operative under the experimental conditions used, at 15 least some label from [1- $^{13}\text{C}_1$] glucose should have been contributed to the zeaxanthin. Furthermore, the mevalonate pathway can at best contribute blocks of two carbon atoms to terpenoids, while in the DXP pathway three carbon units can be delivered to isoprenoids via triose phosphate precursors. Although such three-carbon blocks become separated by the rearrangement involved in the DXP pathway, blocks of three labeled 20 carbon atoms can still be recognized via long-range coupling. Corresponding ^{13}C - ^{13}C long-range couplings have been observed in the biosynthesis of the carotenoid lutein from [2,3,4,5- $^{13}\text{C}_4$] 1-deoxy-D-xylulose by cultured plant cells (*Cantharantus roseus*) [Arigoni et al., Proc. Nat. Acad. Sci. 94:10600-10605 (1997)]. No such long-range coupling was observed in the present experiments with zeaxanthin produced by *Paracoccus* sp. strain 25 R114.

It should be noted that while the results presented here confirm isoprenoid production in *Paracoccus* sp. strain R114 via the mevalonate pathway, and indicate that, under the growth conditions used, there was little or no glucose metabolism via glycolysis, they do not rule out the possibility of some metabolism of glucose via the pentose phosphate pathway in 30 addition to the Entner-Doudoroff pathway. Quantitative determination of glucose metabolism via the latter two pathways could be obtained by analysis of labeling patterns of pyruvate-derived amino acids (as was done for *Paracoccus denitrificans* [Dunstan et al., Biomedical and Environ. Mass Spectrometry 19:369-381 (1990)].

- 76 -

Example 4: Cloning and Sequencing of the Genes Encoding IPP Isomerase and the Enzymes of the Mevalonate Pathway from *Paracoccus* sp. strain R114.

- Culture conditions. *Paracoccus* sp. strain R114 was grown at 28°C in F-medium (10 g/l tryptone, 10 g/l yeast extract, 30 g/l NaCl, 10 g/l D-glucose, 5 g/l MgSO₄·7H₂O, pH 7.0) or 5 in the pre-culture medium described in Example 3 above. Liquid cultures were grown in a rotary shaker at 200 rpm.
- Isolation of genomic DNA. A 600-ml culture of *Paracoccus* sp. strain R114 was centrifuged for 10 minutes at 10,000 x g at 4°C and the pellet was washed once with 200 ml lysis buffer (0.1M NaCl, 50mM EDTA, 10mM Tris-HCl, pH 7.5) and once with 100 ml lysis buffer.
- 10 The final pellet was resuspended in 20 ml lysis buffer containing 50 mg lysozyme and 1 mg RNase A (DNase free). After incubation for 15 minutes at 37°C, 1.5 ml of 20% sodium N-lauroyl-sarcosinate and 2.25 mg of proteinase K were added. After incubation at 50°C for 30-60 minutes, the lysate was extracted with one volume of buffer-saturated phenol, pH 7.5-7.8 (LifeTechnologies, Rockville, MD, USA) by gentle but thorough mixing. The 15 emulsion was centrifuged for 20 minutes at 30,000 x g and the aqueous phase was re-extracted with phenol. The phases were separated as before and the aqueous phase was extracted twice with one volume phenol:chloroform (1:1). At this step centrifugation for 20 minutes at 3,200 x g in a swinging bucket rotor was sufficient to obtain satisfactory phase separation. After a final extraction with one volume of chloroform, 0.1 volume 3M 20 sodium-acetate (pH 5.2) was added and the solution was overlaid with 2 volumes ice-cold ethanol. The precipitated DNA was spooled with a glass-rod, soaked in 70% ethanol for 5 minutes, rinsed with chloroform and then air dried for 5-10 minutes. The DNA was re-suspended overnight in 5 ml TE (10mM Tris-HCl, pH 7.5, 1mM EDTA). Since the solution was yellow due to traces of zeaxanthin, the organic extractions and the spooling were 25 repeated as above to obtain a clear preparation.

Isolation of λ-DNA: The Qiagen® Lambda Kit (Qiagen, Hilden, Germany) was used following the manufacturer's instructions.

Polymerase chain reaction (PCR): Oligonucleotides were purchased from LifeTechnologies (Rockville, MD, USA). PCR was performed in a GeneAmp® PCR system 9700 (PE 30 Applied Biosystems, Foster City, CA, USA) using the GC-rich PCR system (Roche Molecular Biochemicals, Mannheim, Germany) according to the manufacturers instructions. Typically, the MgCl₂ concentration used was 1.5mM and the resolution solution was added to 1M final concentration.

DNA Labeling and detection: The PCR DIG Probe Synthesis Kit and the DIG 35 Luminescent Detection Kit were used for DNA labeling and detection, respectively (both obtained from Roche Molecular Biochemicals, Mannheim, Germany)

DNA sequencing: Sequencing reactions were performed using the BigDye[®] DNA sequencing kit (PE Applied Biosystems, Foster City, CA, USA) according to the manufacturers instructions. Sequencing reactions were purified on DyeExTM spin columns (Qiagen, Hilden, Germany) and fragment separation and detection was done with an ABI PrismTM 310 Genetic Analyzer (PE Applied Biosystems, Foster City, CA, USA).

5 λ -library: A custom made library with partially *Sau3AI* digested *Paracoccus* sp. strain R114 DNA in lambda FIX[®] II was purchased from Stratagene (La Jolla, CA, USA).

10 Cloning, sequencing and characterization of the mevalonate pathway gene cluster from *Paracoccus* sp. strain R114. One of the enzymes of the mevalonate pathway, mevalonate diphosphate decarboxylase, contains highly conserved regions spanning several amino acids. Three such regions were chosen from an alignment of all available eubacterial mevalonate diphosphate decarboxylases and oligonucleotides were designed using the preferred codon usage found in the carotenoid gene cluster of *Paracoccus* sp. strain R1534 (Table 14).

15 The oligonucleotides designed from two homology regions are shown in Table 15. To reduce the degree of degeneracy, sets of oligonucleotides were designed from each peptide. For instance, oligonucleotides mvd-103a-d differ only in the third nucleotide from the 3' end, each accounting for one possible codon for glycine (GGA, although rarely used, was included because of the close proximity to the 3' end). Alternate amino acids were

20 accounted for by designing oligonucleotides to both residues, e.g. oligonucleotides mvd-101a and mvd-101b are specific for leucine or isoleucine, respectively, in the second position of peptide 1 (Table 15). PCR with oligonucleotides mvd-101 and mvd-104 or mvd-106, using *Paracoccus* sp. strain 114 DNA as template, gave a product of the expected size. The PCR product was cloned in the vector pCR[®]2.1-TOPO (Invitrogen, Carlsbad,

25 CA, USA) and sequenced. The cloned fragment was used as a probe for a Southern analysis of *Paracoccus* sp. strain R114 DNA and was found to hybridize to a *BamHI-SalI* fragment of about 950 bp. *Paracoccus* sp. strain R114 DNA was cut with *BamHI* and *SalI* and the fragments were separated by agarose gel electrophoresis. The region around 950 bp was isolated and cloned in the vector pUC19. This partial library was then screened

30 using the *mvd*-PCR fragment as a probe and the insert of a positive clone was sequenced. In parallel, a λ -library prepared from *Paracoccus* sp. strain R114 DNA was screened using the *mvd*-PCR fragment as a probe. DNA was isolated from two positive λ -clones and cut with *BamHI* and *SalI* or *EcoRI* and *SalI*. A number of the restriction fragments were isolated and cloned in the vector pUC19. Several of the fragments contained sequences homologous to genes encoding proteins of the mevalonate pathway. The clones connecting these individual sequences were obtained by PCR with primers derived from the sequences of

- the cloned restriction fragments using the DNA of the λ -clones as template. The assembled sequence from all fragments (SEQ ID NO:42, 44, 46, 48, 50, and 52) and the sequences of the encoded proteins are shown in the Sequence Listing (SEQ ID Nos:43, 45, 47, 49, 51, and 53). Due to a limitation of the PatentIn Program, operons with overlapping genes cannot be shown as a single sequence. Thus, for each gene in the mevalonate operon, the entire nucleotide sequence of the operon is repeated for each gene. Accordingly, SEQ ID Nos:42, 44, 46, 48, 50, and 52 are identical. For purposes of the present invention, we use SEQ ID NO:42 to refer to the nucleotide sequence of the mevalonate operon.
- 10 The arrangement of the mevalonate pathway genes in the *Paracoccus* sp. strain R114 is unique when compared to known mevalonate gene clusters of other bacteria. Besides *Paracoccus* sp. strain R114, only *Borrelia burgdorferi* and *Streptomyces* sp. strain CL190 (Takagi et al., supra) have all mevalonate genes in a single operon (Wilding et al., supra). In *Streptococcus pyogenes* all mevalonate genes are clustered in a single locus but they are grouped in two operons. All other species have two loci with the two kinases and the mevalonate diphosphate decarboxylase grouped in one operon and the HMG-CoA synthase and the HMG-CoA reductase on a second locus, either forming an operon (in *Streptococcus pneumoniae*) or as separate transcription units. All species except the members of *Staphylococcus* have an additional gene linked with the mevalonate cluster,
- 15 which was recently identified as an IPP isomerase (*idi* gene in *Streptomyces* sp. strain CL190) (Kaneda et al., supra). The two *Enterococcus* species and *Staphylococcus haemolyticus* have an acetyl-CoA acetyltransferase gene linked with the HMG-CoA reductase gene. In the *Enterococcus* species the latter two genes are fused.

- 79 -

Table 14: Codon usage in *Paracoccus* sp. strain R1534 carotenoid (*crt*) genes

Amino acid	Codon	Number used	% Used
A – Ala	GCT	3	1.4
	GCC	96	46.2
	GCA	15	7.2
	GCG	94	45.2
C – Cys	TGT	0	0.0
	TGC	15	100.0
D – Asp	GAT	46	38.0
	GAC	75	62.0
E – Glu	GAA	17	25.4
	GAG	50	74.6
F – Phe	TTT	3	5.6
	TTC	51	94.4
G – Gly	GGT	16	10.8
	GGC	87	58.8
	GGA	5	3.4
	GGG	40	27.0
H – His	CAT	30	56.6
	CAC	23	43.4
I – Ile	ATT	5	6.4
	ATC	72	92.3
	ATA	1	1.3
K – Lys	AAA	4	14.3
	AAG	24	85.7
L – Leu	TTA	0	0.0
	TTG	5	2.9
	CTT	15	8.7
	CTC	11	6.4
	CTA	1	0.6
	CTG	140	81.4
M – Met	ATG	49	100.0
N – Asn	AAT	4	20.0
	AAC	16	80.0

- 80 -

Amino acid	Codon	Number used	% Used
P - Pro	CCT	2	2.3
	CCC	41	47.7
	CCA	3	3.5
	CCG	40	46.5
Q - Gln	CAA	6	11.3
	CAG	47	88.7
R - Arg	CGT	11	7.3
	CGC	103	68.2
	CGA	2	1.3
	CGG	26	17.2
	AGA	2	1.3
	AGG	7	4.6
S - Ser	TCT	1	1.1
	TCC	17	19.5
	TCA	0	0.0
	TCG	39	44.8
	AGT	2	2.3
	AGC	28	32.2
T - Thr	ACT	2	2.7
	ACC	36	48.9
	ACA	4	5.3
	ACG	33	44.0
V - Val	GTT	6	5.7
	GTC	61	57.5
	GTA	1	0.9
	GTG	38	35.8
W - Trp	TGG	27	100.0
Y - Tyr	TAT	28	62.2
	TAC	17	37.8

Table 15: Oligonucleotides designed from two conserved bacterial Mvd peptides.

		SEQ ^
Peptide 1	AlaLeuIleLysTyrTrpGlyLys Ile ²	23
Nucleotide sequence ¹	CCSCTGATCAARTAYTGGGGBAARATC	24
Oligonucleotide mvd-101a (5' – 3')	GCSCTGATCAARTAYTGGGG	25
Oligonucleotide mvd-101b (5' – 3')	GCSATCATCAARTAYTGGGG	26
Oligonucleotide mvd-103a (5' – 3')	ATCAARTAYTGGGTAA	27
Oligonucleotide mvd-103b (5' – 3')	ATCAARTAYTGGGCCAA	28
Oligonucleotide mvd-103c (5' – 3')	ATCAARTAYTGGGGAA	29
Oligonucleotide mvd-103d (5' – 3')	ATCAARTAYTGGGGAAA	30
Peptide 2	ThrMetAspAlaGlyProAsnVal Gln ²	31
Nucleotide sequence ¹ (5'-3')	ACSATGGAYGCGSGBCCSAAYGTS CAR	32
Complement (3'-5')	TGSTACCTRGCSCCVGGSTTRCAS GTY	33
Oligonucleotide mvd-104a (3' – 5')	TGGTACCTACGSCCVGG	34
Oligonucleotide mvd-104b (3' – 5')	TGGTACCTGCGSCCVGG	35
Oligonucleotide mvd-104c (3' – 5')	TGCTACCTACGSCCVGG	36
Oligonucleotide mvd-104d (3' – 5')	TGCTACCTGCGSCCVGG	37
Oligonucleotide mvd-106a (3' – 5')	TACCTACGSCCVGGSTTRCA	38
Oligonucleotide mvd-106b (3' – 5')	TACCTGCGSCCVGGSTTRCA	39
Oligonucleotide mvd-106c (3' – 5')	TACCTACGSCCVGGSGTYCA	40
Oligonucleotide mvd-106d (3' – 5')	TACCTGCGSCCVGGSGTYCA	41

^a: SEQ ID NO:¹ using the preferred codons of *Paracoccus* sp. strain R1534, see table 1² alternate amino acid present in some enzyme

5 S = C or G; R = A or G; Y = C or T; B = C or G or T; V = A or C or G

The genes of the mevalonate operon from *Paracoccus* sp. strain R114 were identified by homology of the gene products to proteins in general databases. An amino acid alignment of the HMG-CoA reductase from *Paracoccus* sp. strain R114 (SEQ ID NO:43) was performed with bacterial class I HMG-CoA reductases of *Streptomyces* sp. Strain CL190 (SEQ

- 82 -

- ID NO:54), *S. griseolosporeus* (SEQ ID NO:55), and *Streptomyces* sp. strain KO-3899 (SEQ ID NO:56). EMBL/GenBank/DDBJ database accession numbers are q9z9n4 for *Streptomyces* sp. strain CL190, q9zhn1 for *S. griseolosporeus* and q9zhn0 for *Streptomyces* sp. strain KO-3899. There are two classes of HMG-CoA reductases [Bochar et al., Mol. Genet.
- 5 Metab. 66:122-127 (1999); Boucher et al., Mol. Microbiol. 37:703-716 (2000)]. Eubacterial HMG-CoA reductases are generally of class II, whereas class I enzymes are found in eukaryotes and archaea. The *Streptomyces* and the *Paracoccus* HMG-CoA reductases together with the enzyme from *Vibrio cholerae* are the only eubacterial HMG-CoA reductases of class I known so far.
- 10 An amino acid alignment of isopentenyl diphosphate isomerase (IPP isomerase) (*idi*) from *Paracoccus* sp. strain R114 (SEQ ID NO:45) was performed with close homologs found in the EMBL database, i.e. *Erwinia herbicola* (Q01335) (SEQ ID NO:57), *Borrelia burgdorferi* (O51627) (SEQ ID NO:58), *Synechocystis* sp. PCC 6803 (P74287) (SEQ ID NO:59), *Streptomyces* sp. CL190 (Q9KWF6) (SEQ ID NO:60), *Streptomyces griseolosporeus* (Q9KWF6) (SEQ ID NO:61), *Sulfolobus solfataricus* (P95997) (SEQ ID NO:62), *Rickettsia prowazekii* (Q9ZD90) (SEQ ID NO:63), *Deinococcus radiodurans* (Q9RVE2) (SEQ ID NO:64), *Aeropyrum pernix* (Q9YB30) (SEQ ID NO:65), *Halobacterium* sp. NRC-1 (O54623) (SEQ ID NO:66), *Archaeoglobus fulgidus* (O27997) (SEQ ID NO:67), *Pyrococcus abyssi* (Q9UZS9) (SEQ ID NO:68), *Pyrococcus horikoshii* (O58893) (SEQ ID NO:69),
- 15 20 *Methanobacterium thermoautotrophicum* (O26154) (SEQ ID NO:70), *Methanococcus jannaschii* (Q58272) (SEQ ID NO:71), *Thermoplasma acidophilum* (CAC11250) (SEQ ID NO:72) and *Leishmania major* (Q9NDJ5) (SEQ ID NO:73). EMBL/GenBank/DDBJ database accession numbers are given after the organism's name in parentheses. The first nine sequences are from eubacteria and the next eight sequences are from archaea. Interestingly,
- 25 30 one eukaryotic species, the protozoan parasite *Leishmania major* (SEQ ID NO:73), also has a protein that is highly homologous. This is unexpected because other eukaryotes have a different *idi*, designated type 1 (Kaneda et al., supra). A conserved hypothetical protein from *Bacillus subtilis*, YpgA, also has substantial homology but is considerably smaller than the type 2 *idi*'s.
- 35 An amino acid alignment of bacterial HMG-CoA synthase from *Paracoccus* sp. strain R114 (SEQ ID NO:47) was performed with close homologs found in the EMBL database, i.e. *Streptococcus pneumoniae* (AAG02453) (SEQ ID NO:74), *Streptococcus pyogenes* (AAG02448) (SEQ ID NO:75), *Enterococcus faecalis* (AAG02438) (SEQ ID NO:76), *Enterococcus faecium* (AAG02443) (SEQ ID NO:77), *Staphylococcus haemolyticus* (AAG02427) (SEQ ID NO:78), *Staphylococcus epidermidis* (AAG02433) (SEQ ID NO:79),

- 83 -

Staphylococcus aureus (AAG02422) (SEQ ID NO:80), *Staphylococcus carnosus* (Q9ZB67) (SEQ ID NO:81), *Streptomyces* sp. CL190 (Q9KWF1) (SEQ ID NO:82), *Streptomyces griseolosporeus* (Q9KWF5) (SEQ ID NO:83) and *Borrelia burgdorferi* (051626) (SEQ ID NO:84). EMBL/GenBank/DDBJ database accession numbers are given after each 5 organism's name in parentheses. The first 43 amino acids of the sequence from *Streptomyces griseolosporeus* are missing in the database version.

An amino acid alignment of bacterial mevalonate diphosphate decarboxylase from *Paracoccus* sp. strain R114 (SEQ ID NO:53) was performed with the orthologous proteins from other bacteria, i.e. *Streptococcus pneumoniae* (AAG02456) (SEQ ID NO:85), *Streptococcus pyogenes* (AAG02451) (SEQ ID NO:86), *Enterococcus faecalis* (AAG02441) (SEQ ID NO:87), *Enterococcus faecium* (AAG02446) (SEQ ID NO:88), *Staphylococcus haemolyticus* (AAG02431) (SEQ ID NO:89), *Staphylococcus epidermidis* (AAG02436) (SEQ ID NO:90), *Staphylococcus aureus* (AAG02425) (SEQ ID NO:91), *Streptomyces* sp. CL190 (Q9KWF4) (SEQ ID NO:92), *Streptomyces griseolosporeus* (Q9KWF8) (SEQ ID NO:93) and *Borrelia burgdorferi* (051629) (SEQ ID NO:94). EMBL/GenBank/DDBJ database accession numbers 10 are given after each organism's name in parentheses. 15

Two proteins from *Myxococcus xanthus*, Tac and Taf (database accession numbers q9xb06 and q9xb03, respectively) and a protein from *B. subtilis*, PksG, a putative polyketide biosynthesis protein (database accession number p40830), have substantial homology to the 20 *Paracoccus* sp. strain R114 HMG-CoA synthase. The homology between the *Paracoccus* sp. strain R114 HMG-CoA synthase and the Tac and Taf proteins of the *M. xanthus* is greater than the homology between the HMG-CoA synthases from *Paracoccus* sp. strain R114 and eukaryotes. The bacterial HMG-CoA synthases and the bacterial mevalonate diphosphate decarboxylases share substantial homology with their eukaryotic orthologs. Archaeal 25 HMG-CoA synthases form a more distantly related group of enzymes (Wilding et al., supra) and no mevalonate diphosphate decarboxylase orthologs are found in archaea [Smit and Mushegian, Genome Res. 10:1468-1484 (2000)].

Alignments of the mevalonate kinase (Mvk) (SEQ ID NO:49) and the phosphomevalonate kinase (Pmk) (SEQ ID NO:51) from *Paracoccus* sp. strain R114 were performed to the 30 orthologous proteins from other bacteria, i.e. *Streptococcus pneumoniae* (AAG02455) (SEQ ID NO:95), *Streptococcus pyogenes* (AAG02450) (SEQ ID NO:96), *Enterococcus faecalis* (AAG02440) (SEQ ID NO:97), *Enterococcus faecium* (AAG02445) (SEQ ID NO:98), *Staphylococcus haemolyticus* (AAG02430) (SEQ ID NO:99), *Staphylococcus epidermidis* (AAG02435) (SEQ ID NO:100), *Staphylococcus aureus* (AAG02424) (SEQ ID NO:101), 35 *Streptomyces* sp. CL190 (Q9KWF5) (SEQ ID NO:102), *Streptomyces griseolosporeus*

- 84 -

- (Q9KWF9) (SEQ ID NO:103) and *Borrelia burgdorferi* (051631) (SEQ ID NO:104)(Mvk); and *Streptococcus pneumoniae* (AAG02457) (SEQ ID NO:105), *Streptococcus pyrogenes* (AAG02452) (SEQ ID NO:106), *Enterococcus faecalis* (AAG02442) (SEQ ID NO:107), *Enterococcus faecium* (AAG02447) (SEQ ID NO:108), *Staphylococcus haemolyticus* (AAG02432) (SEQ ID NO:109), *Staphylococcus epidermidis* (AAG02437) (SEQ ID NO:110), *Staphylococcus aureus* (AAG02426) (SEQ ID NO:111), *Streptomyces* sp. CL190 (Q9KWF3) (SEQ ID NO:112), *Streptomyces griseolosporeus* (Q9KWF7) (SEQ ID NO:113) and *Borrelia burgdorferi* (051630) (SEQ ID NO:114) (Pmk). EMBL/GenBank/DDBJ database accession numbers are given after each organism's name in parentheses.
- 10 There is much less homology among the bacterial kinases than among the bacterial orthologs of the other enzymes of the mevalonate pathway. The mevalonate kinase from *Paracoccus* sp. strain R114 (SEQ ID NO:49) has a 37 amino acid insert in the amino-terminal region, which is lacking in other mevalonate kinases. Together with the bacterial Mvk's some archaeal enzymes, e.g. from *Archaeoglobus fulgidus*, *Methanobacterium thermoautotrophicum* and *Pyrococcus abyssi*, are among the best homologues to the Mvk from *Paracoccus* sp. strain R114. The homology among bacterial phosphomevalonate kinases is even weaker than the homology among the bacterial mevalonate kinases. The proteins with the best homologies to the Pmk from *Paracoccus* sp. strain R114 (SEQ ID NO:51) are Mvk's from archaea, e.g. *Aeropyrum pernix*, *Pyrococcus horikoshii*, *M. thermoautotrophicum*, *P. abyssi* and *A. fulgidus*. Since no Pmk's are found in archaea (Smit and Mushegian, supra), this suggests that the same kinase might perform both phosphorylations.

Example 5: Over-expression of the Mevalonate Pathway Genes and the *idi* Gene from *Paracoccus* sp. strain R114 in *E. coli*

- Cloning and expression of the mevalonate operon in *E. coli*. A λ clone, designated clone 25 16, from the *Paracoccus* sp. strain R114 λ library (see Example 4) was used as a template for PCR amplification of the entire mevalonate operon. Primers Mevop-2020 and Mevop-9027 (Table 16) were used for PCR.

Table 16. Primers used for amplification of mevalonate operon from *Paracoccus* sp. strain R114.

Primer	Sequence (5'→3')
Mevop-2020	GGGCAAGCTTGTCCACGGCACGACCAAGCA (SEQ ID NO:115)
Mevop-9027	CGTAATCCGGGCCGCGTTCCAGCGCGTC (SEQ ID NO:116)

The resulting PCR product was cloned in TOPO-XL (Invitrogen, Carlsbad, CA, USA), resulting in plasmid TOPO-XL-mev-op16. The insert carrying the mevalonate operon was excised with *Hind*III and *Sac*I and cloned in the *Hind*III-*Sac*I cut vector pBBR1MCS2 [Kovach et al., Gene 166:175-176 (1995)], resulting in plasmid pBBR-K-mev-op16.

- 5 Plasmid pBBR-K-mev-op16 was used to transform electroporation-competent *E. coli*
strain TG1 [Stratagene, La Jolla, CA; Sambrook et al., In: Nolan, C. (ed.), Molecular
Cloning: A Laboratory Manual (Second Edition), p. A.12 (1989)]. Two representative
positive transformants (*E. coli* TG1/ pBBR-K-mev-op16-1 and *E. coli* TG1/ pBBR-K-mev-
op16-2) were grown in Luria Broth (LB, GibcoBRL, Life Technologies) containing 50 mg/l
10 kanamycin and tested for HMG-CoA reductase activity (encoded by the *Paracoccus* sp.
strain R114 *mvaA* gene) using the methods described in Example 1. *E. coli* does not
possess a gene coding for the enzyme HMG-CoA reductase, hence the lack of detectable
activity. The crude extracts of both representative transformants of *E. coli* TG1/ pBBR-K-
mev-op16 had easily measurable HMG-CoA reductase activity, demonstrating the
15 heterologous expression of the cloned *mvaA* gene.

Table 17. HMG-CoA reductase activity in crude extracts of *E. coli* TG1 cells carrying the cloned mevalonate gene cluster from *Paracoccus* sp. strain R114.

Strain	HMG-CoA reductase activity (U/mg)
<i>E. coli</i> TG1	Not detected ^a
<i>E. coli</i> TG1/ pBBR-K-mev-op16-1	0.25
<i>E. coli</i> TG1/ pBBR-K-mev-op16-2	0.78

^aLess than 0.03 U/mg

- Cloning and expression of the *idi* gene and the individual mevalonate pathway genes from *Paracoccus* sp. strain R114 in *E. coli*. The coding regions of the mevalonate operon genes from *Paracoccus* sp. strain R114 were amplified by PCR using the primers shown in Table 18. The primers were designed such that the ATG start codons constituted the second half of an *Nde*I site (cleavage recognition site CATATG), and *Bam*HI sites (GGATCC) were introduced immediately after the stop codons. All PCR products were cloned in the pCR®2.1-TOPO vector. The names of the resulting vectors are listed in Table 19. Except for the mevalonate kinase gene, all genes contained restriction sites for *Bam*HI, *Nde*I or *Eco*RI, which had to be eliminated in order to facilitate later cloning steps. The sites were eliminated by introducing silent mutations using the QuikChange™ site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA) and the oligonucleotides shown in Table 20.

The mutagenized coding regions were excised from the TOPO-plasmids with *Bam*HI and *Nde*I and ligated with the *Bam*HI-*Nde*I cleaved expression vectors pDS-His and pDS.

- 86 -

These expression vectors were derived from pDSNdeHis, which is described in Example 2 of EP 821,063. The plasmid pDS-His was constructed from pDSNdeHis by deleting a 857 bp *NheI* and *XbaI* fragment carrying a silent chloramphenicol acetyltransferase gene. The plasmid pDS was constructed from pDS-His by replacing a small *EcoRI-BamHI* fragment 5 with the annealed primers S/D-1 (5' AATTAAAGGAGGGTTCATATGAATTG) (SEQ ID NO:117) and S/D-2 (5' GATCCGAATTCATATGAAACCCTCCTT) (SEQ ID NO:118).

Table 18: Oligonucleotides for the cloning of the mevalonate operon genes.

Gene	Forward primer		Reverse primer	
	Name	Sequence (5'-3')	Name	Sequence (5'-3')
<i>mvaA</i>	MvaA-Nde	AAGGCCTCATATGATTTC CCATACCCCGGT (SEQ ID NO:119)	MvaA-Bam	CGGGATCCTCATCGCTCCAT CTCCATGT (SEQ ID NO:120)
<i>idi</i>	Idi-Nde	AAGGCCTCATATGACCGA CAGCAAGGATCA (SEQ ID NO:121)	Idi-Bam	CGGGATCCTCATTGACGGAT AAGCGAGG (SEQ ID NO:122)
<i>hsc</i>	Hcs-Nde	AAGGCCTCATATGAAAGT GCCTAAGATGA (SEQ ID NO:123)	Hcs-Bam	CGGGATCCTCAGGCCTGCCG GTCGACAT (SEQ ID NO:124)
<i>mvk</i>	Mvk-Nde ¹	AAGGCCTCATATGAGCAC CGGCAGGCCTGAAGCA (SEQ ID NO:125)	Mvk-Bam ²	CGGGATCCTCATCCCTGCC CGGCAGCGGTT (SEQ ID NO:126)
<i>pmk</i>	Pmk-Nde	AAGGCCTCATATGGATCA GGTCATCCCGCG (SEQ ID NO:127)	Pmk-Bam	CGGGATCCTCAGTCATCGAA AACAAAGTC (SEQ ID NO:128)
<i>mvd</i>	Mvd-Nde	AAGGCCTCATATGACTGA TGCCGTCCCGCGA (SEQ ID NO:129)	Mvd-Bam	CGGGATCCTCAACGCCCTC GAACGGCG (SEQ ID NO:130)

¹The second codon TCA was changed to AGC (silent mutation - both codons encode

10 serine).

²The last codon GGC was changed to GGA (silent mutation - both codons encode glycine).

- 87 -

Table 19: Names of expression plasmids and construction intermediates.

Gene	PCR fragments in pCR® 2.1-TOPO	After first mutagenesis	After 2 nd mutagenesis	Genes in pDS	Genes in pDS-His
<i>mvaA</i>	TOPO- <i>mvaA</i> -BB	TOPO- <i>mvaA</i> -B	TOPO- <i>mvaA</i>	pDS- <i>mvaA</i>	pDS-His- <i>mvaA</i>
<i>idi</i>	TOPO-ORFX-B	TOPO- <i>idi</i>	n/a	pDS- <i>idi</i>	pDS-His- <i>idi</i>
<i>hsc</i>	TOPO- <i>hcs</i> -EN	TOPO- <i>hcs</i> -N	TOPO- <i>hcs</i>	pDS- <i>hcs</i>	pDS-His- <i>hcs</i>
<i>mvk</i>	TOPO- <i>mvk</i>	n/a	n/a	pDS- <i>mvk</i>	pDS-His- <i>mvk</i>
<i>pmk</i>	TOPO- <i>pmk</i> -B	TOPO- <i>pmk</i>	n/a	Nd	pDS-His- <i>pmk</i>
<i>mvd</i>	TOPO- <i>mvd</i> -B	TOPO- <i>mvd</i>	n/a	pDS- <i>mvd</i>	pDS-His- <i>mvd</i>

n/a: not applicable; nd: not done

Table 20: Oligonucleotides for site-directed mutagenesis.

Gene/ Site	Forward primer		Reverse primer	
	Name	Sequence (5'-3')	Name	Sequence (5'-3')
<i>mvaA</i> / <i>BamHI</i> -1	Mva-B1up	CCGGCATT CGGGCGGC ATCCAGGTCTCGCTG (SEQ ID NO:131)	Mva-B1down	CAGCGAGACCTGGATG CCGCCC GAATGCCGG (SEQ ID NO:132)
<i>mvaA</i> / <i>BamHI</i> -2	Mva-B2up	CGTGCAGGGCTGGATT CTGTCGGAATACCCG (SEQ ID NO:133)	Mva-B2down	CGGGTATTCCGACAGA ATCCAGCCCTGCACG (SEQ ID NO:134)
<i>idi</i> / <i>BamHI</i>	Idi-Bup2	GGGCTGCGCGCCGGCA TCCGGCATT CGACG (SEQ ID NO:135)	Idi-Bdown2	CGTCGAAATGCCGGAT GCCGGCGCGCAGCCC (SEQ ID NO:136)
<i>hcs</i> / <i>EcoRI</i>	Hcs-Eup	GGGTGCGACGGGCGAG TTCTTCGATGCGCGG (SEQ ID NO:137)	Hcs-Edown	CCGCGCATCGAAGAAC TCGCCCCGTCGCACCC (SEQ ID NO:138)
<i>hcs</i> / <i>NdeI</i>	Hcs-Nup-c	CACGCCCGTCACATAC GACGAATA CGTTGCC (SEQ ID NO:139)	Hcs-Ndown-c	GGCAACGTATT CGTCG TATGTGACGGGCGTG (SEQ ID NO:140)
<i>pmk</i> / <i>BamHI</i>	Pmk-Bup	GAGGCTCGGGCTGGC TCCTCGGCGGCCGTG (SEQ ID NO:141)	Pmk-Bdown	CACCGCCGCCGAGGAG CCAAGCCCGAGCCTC (SEQ ID NO:142)
<i>mvd</i> / <i>BamHI</i>	Mvd-Bup	CGGCACGCTGCTGGAC CCGGGCGACGCCCTC (SEQ ID NO:143)	Mvd-Bdown	GAAGGCGTCGCCGGG TCCAGCAGCGTGCCG (SEQ ID NO:144)

- 5 *E. coli* strain M15 [Villarejo and Zabin, J. Bacteriol. 120:466-474 (1974)] carrying the *lacI* (lac repressor)-containing plasmid pREP4 (EMBL/GenBank accession number A25856) was transformed with the ligation mixtures and recombinant cells were selected for by

growth on LB-Agar plates supplemented with 100 mg/L ampicillin and 25 mg/L kanamycin. Positive clones containing the correct mevalonate operon gene insert were verified by PCR.

- For expression of the inserted genes, each of the *E. coli* strains were grown overnight at 37°C in LB medium containing 25 mg/L kanamycin and 100 mg/L ampicillin. The next day, 25 ml of fresh medium was inoculated with 0.5 ml of the overnight cultures and the new cultures were grown at 37°C. When the OD₆₀₀ of the cultures reached 0.4, expression of the cloned genes was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM, and the incubation of the cultures (with shaking) was continued for four hours, after which the cells were collected by centrifugation.

Crude extract preparation, HMG-CoA reductase assays, and IPP isomerase assays were performed as described in Example 1. Tables 21 and 22 show the HMG-CoA reductase and IPP isomerase activities, respectively, in the recombinant *E. coli* strains. Upon IPTG induction, strains M15/pDS-*mvaA* and M15/pDS-*idi* contained high levels of the HMG-CoA reductase and IPP isomerase activity, respectively. This illustrates the ability to over-express the mevalonate pathway genes (and overproduce their cognate gene products in an active form) from *Paracoccus* sp. strain R114 in *E. coli*.

Table 21. Induction of HMG-CoA reductase activity in *E. coli* strains over-expressing the cloned *mvaA* gene from *Paracoccus* sp. strain R114.

Strain/plasmid	IPTG Induction	HMG-CoA reductase activity (U/mg)
M15/pDS- <i>mvaA</i>	-	8.34
M15/pDS- <i>mvaA</i>	+	90.0
M15/pDS-His- <i>mvaA</i>	-	1.74
M15/pDS-His- <i>mvaA</i>	+	2.95
M15/pDS- <i>mvd</i> ^a	-	0.05

20 ^aM15/pDS-*mvd* was included as a negative control

- 89 -

Table 22. Induction of IPP isomerase activity in *E. coli* strains over-expressing the cloned *idi* gene from *Paracoccus* sp. strain R114.

Strain/plasmid	IPTG Induction	IPP isomerase activity (U/mg)
M15/pDS- <i>idi</i>	-	Not detected ^b
M15/pDS- <i>idi</i>	+	22.0
M15/pDS-His- <i>idi</i>	-	Not detected
M15/pDS-His- <i>idi</i>	+	Not detected
M15/pDS- <i>mvd</i> ^a	-	Not detected

^aM15/pDS-*mvd* was included as a negative control

^b<1 U/mg

- 5 The crude extracts used for the enzyme assays were analyzed by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE). For strains *E. coli* M15/pDS-*mvaA* and *E. coli* M15/pDS-His-*mvaA*, the presence or absence of a highly expressed protein of the expected molecular mass (36.3 kD) correlated with the HMG-CoA reductase activity measured in the extracts (Table 21). The absence of the His-tagged protein could be
10 explained by reduced expression at the level of transcription or translation by instability of the mRNA or the protein. The crude extracts of *E. coli* M15/pDS-*idi* and *E. coli* M15/pDS-His-*idi* both showed highly expressed proteins of the expected molecular masses of 37.3 kD and 39.0 kD, respectively. However, only the extract from *E. coli* M15/pDS-*idi* had increased IPP isomerase activity (Table 22), indicating that the histidine-tagged form of
15 the enzyme was not functional under these conditions.

- By SDS-PAGE analysis of crude extracts of *E. coli* strains over-expressing the other four genes of the *Paracoccus* sp. strain R114 mevalonate operon (*hcs*, *pmk*, *mvk*, and *mvd*, refer to Table 19) high expression of the native form of the enzyme was not detected upon IPTG induction, although some expression cannot be ruled out. On the other hand, high
20 expression was observed with the His-tagged form of all four proteins.

Example 6: Improved Zeaxanthin Production in *Paracoccus* sp. strain R114 by Over-Expression of the *crtE* Gene

Construction of pBBR-K-Zea4, pBBR-K-Zea4-up and pBBR-K-Zea4-down and effects of these plasmids on zeaxanthin production in *Paracoccus* sp. strain R114. The carotenoid (*crt*) gene cluster of *Paracoccus* sp. strain R1534 was excised from plasmid pZea-4 [Pasanmontes et al., Gene 185:35-41 (1997)] as an 8.3 kb *Bam*HI - *Eco*RI fragment. This fragment containing the *crt* gene cluster was ligated into the *Bam*HI and *Eco*RI-cut vector pBBR1MCS-2 (GenBank accession #U23751) resulting in pBBR-K-Zea4. Plasmid pBBR-

- 90 -

K-Zea4 was introduced into *Paracoccus* sp. strain R114 by conjugation to test for improved zeaxanthin production. The control strain R114 and two independent isolates of strain R114/pBBR-K-Zea4 were tested for zeaxanthin production in shake flask cultures (using medium 362F/2, see Example 11). The data in Table 23 show that both recombinant strains carrying plasmid pBBR-K-Zea4 produced significantly higher levels of zeaxanthin than R114, and had higher specific rates of production (mg zeaxanthin/OD₆₆₀). This suggested that one or more of the genes within the cloned insert in pBBR-K-Zea4 encoded an enzyme(s) that was limiting for zeaxanthin production in *Paracoccus* sp. strain R114.

Table 23. Zeaxanthin production by strains R114 and R114/pBBR-K-Zea4.

Strain	24 Hours		48 Hours		72 Hours	
	ZXN ^a (mg/l)	Spec. Form. ^b	ZXN (mg/l)	Spec. Form.	ZXN (mg/l)	Spec. Form.
R114	54.5	2.2	81.7	4.1	78.1	4.5
R114/pBBR-K-Zea4 (clone 4)	41.0	3.0	100.7	5.2	97.6	6.2
R114/pBBR-K-Zea4 (clone 5)	41.1	3.1	110.5	5.7	102.1	6.5

10 ^aZeaxanthin

^bSpecific Formation (mg ZXN/l/OD₆₆₀)

To localize the positive effect, two plasmid derivatives were created that contained sub-cloned regions of the cloned insert present in pBBR-K-Zea4. The "upstream" region of the pBBR-K-Zea4 insert, comprising ORF 5 and the genes *atoB* and *crtE*, (Pasamontes et al., 15 *supra*) is flanked by unique sites for the restriction enzymes *Xba*I and *Avr*II. Plasmid pBBR-K-Zea4-down was constructed by digesting pBBR-K-Zea4 with these two enzymes and deleting the "upstream" region. Analogously, plasmid pBBR-K-Zea4-up was constructed by deletion of the "downstream" region within the cloned insert in pBBR-K-Zea4, 20 using the restriction enzymes *Eco*RV and *Stu*I. The two new plasmids were transferred to *Paracoccus* sp. strain R114 by conjugation. Zeaxanthin production was compared (shake flask cultures, same conditions as described above) in strains R114 (host control), R114/pBBR-K (empty vector control), R114/ pBBR-K-Zea4-down and R114/pBBR-K-Zea4-up (Table 24). The data clearly showed that the positive effect on zeaxanthin production was a result of the presence in multiple copies of the cloned segment containing 25 ORF5, *atoB* and *crtE*, i.e., the insert present in plasmid pBBR-K-Zea4-up. A series of deletion plasmids was constructed from pBBR-K-Zea4-up. By introducing each of these plasmids into strain R114 and testing for zeaxanthin production, it was determined that it

- 91 -

was over-expression of the *crtE* gene that was providing the improved zeaxanthin production in strains R114/pBBR-K-Zea4 and pBBR-K-Zea4-up. This result is consistent with the activity of GGPP synthase (encoded by *crtE*) being limiting for zeaxanthin production in *Paracoccus* sp. strain R114. Using the methods described in Example 1, 5 crude extract of strain R114/pBBR-K-Zea4-up was found to have 2.6-fold higher GGPP synthase activity than R114. To prove this directly, a new plasmid allowing over-expression of only the *crtE* gene was constructed as described in the following two sections.

Table 24. Zeaxanthin production by strains carrying deletion derivatives of plasmid pBBR-K-Zea4.

Strain	24 Hours		48 Hours		72 Hours	
	ZXN ^a (mg/l)	Spec. Form. ^b	ZXN (mg/l)	Spec. Form.	ZXN (mg/l)	Spec. Form.
R114	35.0	1.2	75.7	4.1	73.9	4.4
R114/pBBR-K	32.0	1.5	59.3	3.1	63.3	3.9
R114/pBBR-K-Zea4-up	51.5	2.2	98.8	5.5	85.5	5.7
R114/pBBR-K-Zea4-down	41.6	1.8	63.4	3.3	66.4	3.9

10 ^aZeaxanthin

^bSpecific Formation (mg ZXN/l/OD₆₆₀)

Construction of the expression vectors pBBR-K-PcrtE and pBBR-tK-PcrtE. The vector pBBR1MCS-2 was cut with *Bst*XI and *Bsu*36I and the larger fragment was ligated with the annealed oligonucleotides MCS-2 up 15 (5' TCAGAATTCCGGTACCATATGAAGCTTGGATCCGGGG 3') (SEQ ID NO:145) and MCS-2 down (5' GGATCCAAGCTTCATATGGTACCGAATT 3') (SEQ ID NO:146), resulting in vector pBBR-K-Nde. The 270 bp region upstream of the *crtE* gene in the carotenoid gene cluster from *Paracoccus* sp. strain R114, which contains the putative *crtE* promoter (PcrtE) including the ribosome binding site and the *crtE* start codon (Pasamontes et al., supra) was amplified from *Paracoccus* sp. strain R114 DNA by PCR with primers crtE-up (5' GGAATTCTGCTGAACCGATGGCG 3') (SEQ ID NO:147) and crtE-down (5' GGGGTACCATATGTGCCTTCGTTGCGTCAGTC 3') (SEQ ID NO:148). The PCR product was cut with *Eco*RI and *Nde*I and inserted into the *Eco*RI-*Nde*I cut backbone of pBBR-K-Nde, yielding plasmid pBBR-K-PcrtE. An *Nde*I site, which contains the ATG 20 start codon of *crtE*, was included in primer crtE-down. Hence, any introduced coding region with the start codon embedded in a *Nde*I site should be expressed using the ribosomal binding site of *crtE*. The plasmid pBBR-K-PcrtE was cut with *Bam*HI and the 25 annealed oligonucleotides pha-t-up

- 92 -

(5' GATCCGGCGTGTGCGCAATTAAATTGCGCACACGCCCTGCCTTAAAC 3')

(SEQ ID NO:149) and pha-t-down

(5' GATCGTTAACGCAGGGGGCGTGTGCGCAATTAAATTGCGCACACGCCG 3')

(SEQ ID NO:150) were inserted. The insertion was verified by sequencing, and the version

5 of the plasmid having the oligos inserted in the orientation that reconstitutes the *Bam*HI site closer to the *PcrtE* promoter was named pBBR-tK-*PcrtE*. The inserted sequence carries the putative transcriptional terminator found between the *Paracoccus* sp. strain R114 *phaA* and *phaB* genes (see Example 10) and should, therefore, ensure proper termination of the transcripts initiated from the *PcrtE* promoter.

- 10 **Construction of plasmid pBBR-K-*PcrtE-crtE-3*.** To construct a multi-copy plasmid for increased expression of the *crtE* gene in the *Paracoccus* sp. strain R114 host, the *crtE* gene was amplified from plasmid p59-2 (Pasamontes et al., supra) by PCR using the primers *crtE-Nde* (5' AAGGCCTCATATGACGCCAAGCAGCAATT 3') (SEQ ID NO:151) and *crtE-Bam* (5' CGGGATCCTAGGGCGCTGCGGCGGATG 3') (SEQ ID NO:152). The
- 15 amplified fragment was cloned in the pCR® 2.1-TOPO vector, resulting in plasmid TOPO-*crtE*. The *Nde*I-*Bam*HI fragment from TOPO-*crtE* was subcloned in *Nde*I-*Bam*HI-digested plasmid pBBR-K-*PcrtE*, yielding pBBR-K-*PcrtE-crtE*. Finally, pBBR-K-*PcrtE-crtE-3* was constructed by replacing the smaller *Bgl*II fragment from pBBR-K-*PcrtE-crtE* with the smaller *Bgl*II fragment from pBBR-K-Zea4-up. Plasmid pBBR-K-*PcrtE-crtE-3*
- 20 was transferred to *Paracoccus* sp. strain R114 by electroporation. Using the methods described in Example 1, GGPP synthase activity in crude extracts was found to be 2.9-fold higher in strain R114/pBBR-K-*PcrtE-crtE-3* than in strain R114. This degree of elevated activity was similar to that observed in R114/pBBR-K-Zea4-up. Table 25 shows the zeaxanthin production by strain R114/pBBR-K-*PcrtE-crtE-3* was essentially identical to
- 25 strain R114/pBBR-K-Zea4-up.

- 93 -

Table 25. Comparison of zeaxanthin production by strains R114/pBBR-K-PcrtE-crtE-3 and R114/pBBR-K-Zea4-up.

Strain	24 Hours		48 Hours		72 Hours	
	ZXN ^a (mg/l)	Spec. Form. ^b	ZXN (mg/l)	Spec. Form.	ZXN (mg/l)	Spec. Form.
R114	49.0	1.6	83.9	3.3	97.8	4.3
R114/pBBR-K	42.6	1.8	73.7	3.8	88.8	4.9
R114/pBBR-K-PcrtE-crtE-3	64.6	2.9	127.0	5.8	165.6	8.5
R114/pBBR-K-Zea4-up	64.7	2.9	118.0	5.9	158.0	10.1

^aZeaxanthin^bSpecific Formation (mg ZXN/l/OD₆₆₀)

5 **Example 7: Expression of Individual Genes of the *Paracoccus* sp. strain R114 Mevalonate Operon in the Native Host, *Paracoccus* sp. strain R114**

Expression of individual cloned genes of the *Paracoccus* sp. strain R114 mevalonate operon in the *Paracoccus* sp. strain R114 host. The mutagenized coding regions of the mevalonate operon genes in TOPO-plasmids (see Example 5) were excised with *Bam*HI and *Nde*I and

10 ligated with the *Bam*HI-*Nde*I cleaved vector pBBR-tK-PcrtE (see Example 6). The resulting plasmids pBBR-tK-PcrtE-mvaA, pBBR-tK-PcrtE-idi, pBBR-tK-PcrtE-hcs, pBBR-tK-PcrtE-mvk, pBBR-tK-PcrtE-pmk and pBBR-tK-PcrtE-mvd were introduced into *Paracoccus* sp. strain R114 by electroporation. Transformants were selected on agar medium containing 50 mg/l kanamycin and verified by PCR.

15 To illustrate that the plasmid-borne mevalonate pathway genes can be expressed in the native host *Paracoccus* sp. strain R114, HMG-CoA reductase activity was compared in crude extracts of strains R114/pBBR-K (control) and R114/pBBR-tK-PcrtE-mvaA (methods used are set forth in Example 1). The specific activities of HMG-CoA reductase in strains R114/pBBR-K and R114/pBBR-tK-PcrtE-mvaA were 2.37 U/mg and 6.0 U/mg, respectively. Thus the presence of the mvaA gene on a multicopy plasmid (and expressed from the PcrtE promoter) resulted in a 2.5-fold increase in HMG-CoA reductase activity relative to the basal (i.e., chromosomally encoded) activity of R114 carrying the empty vector pBBR-K.

- 94 -

Example 8: Construction of "Mini-Operons" for Simultaneous Over-Expression the Cloned Genes of the Mevalonate Pathway with the *Paracoccus* sp. strain R114 *crtE* Gene

- Plasmid constructions. As was shown in Example 6, introduction of plasmid pBBR-K-5 *PcrtE-crtE-3* into *Paracoccus* sp. strain R114 resulted in increased production of zeaxanthin, indicating that GGPP synthase activity was rate limiting for zeaxanthin biosynthesis in strain R114. Example 7 further showed that genes coding for the enzymes of the mevalonate pathway could be over-expressed in the native host *Paracoccus* sp. strain R114, and resulted in increased activity of the encoded enzyme. However, none of the 10 recombinant strains of *Paracoccus* sp. strain R114 that carried plasmids containing each individual gene of the mevalonate operon showed increased zeaxanthin production compared to strain R114. It is possible that the benefit of over-expression of the genes of the mevalonate operon in *Paracoccus* sp. strain R114 could be masked by the downstream "bottleneck" in the zeaxanthin pathway (GGPP synthase). Creation of plasmids that allow 15 simultaneous over-expression of each mevalonate pathway gene (or perhaps combinations of these genes) together with *crtE* could relieve all rate limitations in the overall zeaxanthin biosynthetic pathway, thereby improving zeaxanthin production. The next section describes the construction of "mini-operons" designed to allow co-over-expression of *crtE* and each of the genes coding for the five enzymes of the mevalonate pathway.
- 20 The *crtE*, *mvaA*, *idi* and *mvk* genes were excised from the respective TOPO-plasmids (described in Examples 5 and 6) with *Bam*HI and *Nde*I and ligated with *Bam*HI-*Nde*I-cleaved vector pOCV-1 (described in Example 12). The *crtE* gene does not have an adenine as the last nucleotide of the coding region, and in addition, has a TAG rather than a TGA stop codon and an unsuitable distance between the stop codon and the *Bam*HI site. 25 Therefore, the end of *crtE* does not meet the requirements of the operon construction vectors (refer to Example 12) and *crtE* must be the last gene in any operon constructed with pOCV-1-*crtE*. To meet the requirement for an adenine as the first nucleotide of the second codon and the last nucleotide of the last codon, mutations had to be introduced in three genes of the mevalonate operon. The second codon of *pmk*, GAT, encoding Asp, was 30 changed into AAT, encoding Asn. The last codon of *mvd* ends with a T and the last codons of *pmk* and *hcs* end with C. Changing these nucleotides to A results in silent mutations except for *pmk* where the last amino acid is changed from Asp to Glu. Oligonucleotides were designed to introduce the necessary changes by PCR. The sequences of the oligonucleotides and the templates used for those PCR reactions are 35 shown in Table 26. All PCR products were cloned in the pCR®2.1-TOPO vector, resulting in plasmids TOPO-*mvd*^{OCV}, TOPO-*pmk*^{OCV} and TOPO-*hcs*^{OCV}. The inserts were excised

- 95 -

with *Nde*I and *Bam*HI and ligated with the *Nde*I-*Bam*HI cut backbone of pOCV-2 (see Example 12). The final cloning steps to assemble each of the "mini-operons" were analogous, and are illustrated by the representative scheme for construction of pBBR-K-*P crtE-mvaA-crtE-3*.

- 5 Table 26: Oligonucleotides and templates used for PCR in the construction of plasmids TOPO-*mvd*^{OCV}, TOPO-*pmk*^{OCV} and TOPO-*hcs*^{OCV}.

Gene	Forward primer		Reverse primer		Template
	Name	Sequence (5'-3')	Name	Sequence (5'-3')	
<i>Hcs</i>	<i>Hcs-</i> <i>Nde</i>	AAGGCCTCATATGAAA GTGCCTAAGATGA (SEQ ID NO: 123)	<i>Hcs-</i> <i>mut3</i>	CCGGATCCTCATGCC TGCCGGTCGACATAG (SEQ ID NO: 153)	pBBR-tK- <i>P crtE-hcs</i>
<i>Pmk</i>	<i>Pmk-</i> <i>mut5</i>	GAAGGCACATATGAAT CAGGTCACTCCGCGC (SEQ ID NO: 154)	<i>Pmk-</i> <i>mut3</i>	GCCGGATCCTCATTC ATCGAAAACAAGTCC (SEQ ID NO: 155)	pBBR-tK- <i>P crtE-pmk</i>
<i>Mvd</i>	<i>Mvd-</i> <i>Nde</i>	AAGGCCTCATATGACT GATGCCGTCCGCGA (SEQ ID NO: 129)	<i>Mvd-</i> <i>mut3</i>	ACGCCGGATCCTCAT CGCCCTCGAACGGC (SEQ ID NO: 156)	pBBR-tK- <i>P crtE-mvd</i>

Example 9: Cloning and Sequencing of the *ispA* Gene Encoding FPP Synthase from *Paracoccus* sp. strain R114

- 10 Because FPP synthase lies in the central pathway for zeaxanthin biosynthesis in *Paracoccus* sp. strain R114, increasing the activity of this enzyme by increasing the dosage of the *ispA* gene has the potential to improve zeaxanthin production. For this reason, the *ispA* gene from *Paracoccus* sp. strain R114 was cloned and sequenced as follows. The amino acid sequences of six bacterial FPP synthases were obtained from public databases. These
- 15 sequences have several highly conserved regions. Two such regions, and the oligonucleotides used for PCR, are shown in Table 27. PCR with oligonucleotides GTT-1 and GTT-2, using *Paracoccus* sp. strain R114 DNA as template, gave a product of the expected size. The PCR product was cloned in the vector pCR® 2.1-TOPO and sequenced. The cloned fragment was used as a probe for a Southern analysis of *Paracoccus* sp. strain R114 DNA
- 20 and was found to hybridize to a *Bam*HI-*Nco*I fragment of about 1.9 kb. *Paracoccus* sp. strain R114 DNA was cut with *Bam*HI and *Nco*I and the fragments were separated by agarose gel electrophoresis. The region between 1.5 and 2.1 kb was isolated and cloned in the *Bam*HI and *Nco*I sites of a cloning vector. This partial library was then screened using the *ispA*-PCR fragment as a probe, and two positive clones were isolated. Sequencing confirmed that the plasmids of both clones contained the *ispA* gene. Upstream of *ispA* (SEQ ID NO:159) is the gene for the small subunit of exonuclease VII, XseB (SEQ ID NO:158),

and downstream is the *dxs* gene (SEQ ID NO:160) encoding the 1-deoxyxylulose-5-phosphate synthase. This is the same gene arrangement as found in *E. coli*. The sequence of the *NcoI-BamHI* fragment is illustrated in SEQ ID NO:157, the amino acid sequences of XseB, IspA and Dxs are illustrated in SEQ ID NO:158, SEQ ID NO:159, and SEQ ID NO:160, respectively. The start codon of *ispA* may be GTG or ATG resulting in two or one methionine residues, respectively, at the amino-terminus of the native IspA.

Using the same general cloning strategy described in Examples 5-7, a new plasmid, pBBR-tK-P crtE-*ispA*-2 was constructed to allow for over-expression of the *ispA* gene in the native host *Paracoccus* sp. strain R114. The plasmid was introduced into strain R114 by electroporation, and transformants were confirmed by PCR. Three representative transformants and a control strain (R114/pBBR-K) were grown in 362F/2 medium (Example 11), crude extracts were prepared and assayed for activity of the *ispA* gene product, FPP synthase according to the methods described in Example 1. The basal (chromosomally-encoded) FPP synthase specific activity in R114/pBBR-K was 62.6 U/mg. The FPP synthase activity in the three transformants was 108.3 U/mg (73% increase), 98.5 U/mg (57% increase) and 83.8 U/mg (34% increase), demonstrating the over-expression of the *ispA* gene and over-production of its product, FPP synthase, in an active form in *Paracoccus* sp. strain R114.

Table 27: Oligonucleotides designed from two conserved bacterial IspA peptides.

<u>Peptide 1</u>	
<i>Bradyrhizobium japonicum</i>	Val His Asp Asp Leu Pro (SEQ ID NO:161)
<i>Rhizobium</i> sp. strain NGR234	Val His Asp Asp Leu Pro (SEQ ID NO:162)
<i>Bacillus stearothermophilus</i>	Ile His Asp Asp Leu Pro (SEQ ID NO:163)
<i>Bacillus subtilis</i>	Ile His Asp Asp Leu Pro (SEQ ID NO:164)
<i>Escherichia coli</i>	Ile His Asp Asp Leu Pro (SEQ ID NO:165)
<i>Haemophilus influenzae</i>	Ile His Asp Asp Leu Pro (SEQ ID NO:166)
Oligonucleotide GTT-1 (5'-3')	tc cay gay gay ctg cc (SEQ ID NO:167)
<u>Peptide 2</u>	
<i>Bradyrhizobium japonicum</i>	Asp Asp Ile Leu Asp (SEQ ID NO:168)
<i>Rhizobium</i> sp. strain NGR234	Asp Asp Ile Leu Asp (SEQ ID NO:169)
<i>Bacillus stearothermophilus</i>	Asp Asp Ile Leu Asp (SEQ ID NO:170)
<i>Bacillus subtilis</i>	Asp Asp Ile Leu Asp (SEQ ID NO:171)
<i>Escherichia coli</i>	Asp Asp Ile Leu Asp (SEQ ID NO:172)
<i>Haemophilus influenzae</i>	Asp Asp Ile Leu Asp (SEQ ID NO:173)
Reverse complement of Oligonucleotide GTT-2 (5'-3')	gay gay atc ctg gay (SEQ ID NO:174)

Y = C or T

**Example 10: Cloning and Sequencing of the Genes Coding for Acetyl-CoA
Acetyltransferase from *Paracoccus* sp. strain R114**

The first committed step in IPP biosynthesis is the condensation of acetyl-CoA and acetoacetyl-CoA to hydroxymethylglutaryl-CoA (HMG-CoA) by HMG-CoA synthase. The substrate acetoacetyl-CoA is formed by the enzyme acetyl-CoA acetyltransferase (also known as acetoacetyl-CoA thiolase or β -ketothiolase) by condensation of two molecules of acetyl-CoA. Because this reaction links central metabolism (at acetyl-CoA) to isoprenoid biosynthesis via the mevalonate pathway, increasing the activity of acetyl-CoA acetyltransferase by gene amplification has the potential to increase carbon flow to carotenoids and other isoprenoids *in vivo*. In *Paracoccus* sp. strain R114, there are at least two genes, *atoB* and *phaA*, that encode acetyl-CoA acetyltransferases. The end of the *atoB* gene is 165 nucleotides upstream of the start of *crtE* in *Paracoccus* sp. strains R1534 (US 6,087,152) and R114 (this work). The nucleotide sequence of the *atoB* gene and the corresponding amino acid sequence of the encoded acetyl-CoA acetyltransferase from *Paracoccus* sp. strain R1534 are illustrated in SEQ ID NO:175 and SEQ ID NO:176, respectively.

Using the same general strategy as described in Example 5, the *atoB* gene was cloned in plasmids pDS and pDS-His. The new plasmids, pDS-*atoB* and pDS-His-*atoB* were introduced into *E. coli* strain M15. The resulting strains M15/pDS-*atoB* and M15/pDS-His-*atoB* were grown with and without IPTG induction (as described in Example 5), and crude extracts were prepared for acetyl-CoA acetyltransferase assays (methods used were described in Example 1) and SDS-PAGE analysis. The acetyl-CoA acetyltransferase specific activities in extracts of M15/pDS-*atoB* and M15/pDS-His-*atoB* (with IPTG induction) were 0.2 U/mg and 13.52 U/mg, respectively. The basal activity measured in *E. coli* without the plasmids was 0.006 U/mg. Upon IPTG induction the *atoB* gene product, acetyl-CoA acetyltransferase, is overproduced in *E. coli* M15. Both the native (M15/pDS-*atoB*) and His-tagged (M15/pDS/his-*atoB*) forms were overproduced. The degree of overproduction was much higher in M15/pDS-His-*atoB*, consistent with the measured acetyl-CoA acetyltransferase activity in the (induced) extracts of the two strains.

Acetoacetyl-CoA is also the substrate for the first committed step in polyhydroxyalkanoate (PHA) biosynthesis. In many bacteria the genes involved in PHA biosynthesis are grouped in operons [Madison and Huisman, *Microbiol. Mol. Biol. Rev.*, 63:21-53 (1999)]. In *Paracoccus denitrificans* the *phaA* and *phaB* genes, encoding the acetyl-CoA acetyltransferase and acetoacetyl-CoA reductase, respectively, are clustered in

an operon [Yabutani et al., FEMS Microbiol. Lett. 133:85-90 (1995)] whereas *phaC*, the gene encoding the last enzyme in the pathway, poly(3-hydroxyalkanoate) synthase, is not part of this operon [Ueda et al., J. Bacteriol. 178:774-779 (1995)]. PCR fragments containing parts of *phaA* from *Paracoccus* sp. strain R1534 and *phaC* from *Paracoccus* sp. strain R114 were obtained using primers based on the *P. denitrificans* *phaA* and *phaC* gene sequences. The PCR fragments were then used as probes to screen a *Paracoccus* sp. strain R114 λ-library (see Example 4). Several λ-clones hybridizing with the *phaA* or the *phaC* probes were isolated, and the presence of the *phaA* or *phaC* genes in the inserts was verified by sequence analysis. Three *phaA* λ-clones were further analyzed by subcloning and sequencing, whereby the *phaB* was found downstream of *phaA*. Therefore, as is the case in *P. denitrificans*, the *phaA* and *phaB* genes are clustered whereas the *phaC* gene is located elsewhere in the genome. The nucleotide sequence of the *phaAB* cluster from *Paracoccus* sp. strain R114 and the deduced amino acid sequences of the acetyl-CoA acetyltransferase (PhaA) are illustrated in SEQ ID NO:177, and SEQ ID NOs:178 and 179, respectively. The clustering of genes involved in PHA biosynthesis in operons suggests that at least *phaA* and *phaB* are expressed together when the cell produces poly(3-hydroxyalkanoates). On the other hand, a putative transcriptional stop signal is found between the *phaA* and *phaB* genes from *Paracoccus* sp. strain R114 which is absent in the *P. denitrificans* *phaAB* operon (Yabutani et al., supra). Thus, the expression of the two genes might not be coupled in *Paracoccus* sp. strain R114.

Using the same general strategy as described in Example 5, the *phaA* gene was cloned in plasmid pDS-His. The new plasmid, pDS-His-*phaA*, was introduced into *E. coli* strain M15. The resulting strain M15/pDS-His-*phaA* was grown with and without IPTG induction (as described in Example 5) and crude extracts were prepared for SDS-PAGE analysis. The cloned His-tagged *Paracoccus* sp. strain R114 PhaA (acetyl-CoA acetyl-transferase) is overproduced upon IPTG induction in the *E. coli* M15 host.

The potential benefit of amplifying the *atoB* or *phaA* genes, encoding acetyl-CoA acetyl-transferase, on zeaxanthin production is mentioned above. In addition, it may be beneficial for zeaxanthin production to decrease or eliminate the activity of actoacetyl-CoA reductase (the *phaB* gene product) to avoid diversion of some of the acetoacetyl-CoA formed *in vivo* to the PHA pathway. Mutants of *Paracoccus* sp. strain R114 lacking activity of *phaB* could be obtained by gene replacement techniques (specifically replacing the wild-type *phaB* gene in the chromosome with an inactive form of the gene) or by classical mutagenesis and screening.

Example 11: Model for the Industrial Production of Lycopene Using Mutants Derived from *Paracoccus* sp. strain R114

Lycopene is a red carotenoid that is an intermediate in the biosynthesis of zeaxanthin in the new *Paracoccus* species represented by strain R-1512 and its mutant derivatives R1534 5 and R114. As lycopene itself has significant commercial potential, it was of interest to test the potential of the new *Paracoccus* species to produce lycopene by industrial fermentation. To obtain mutants blocked in zeaxanthin biosynthesis that accumulated lycopene, *Paracoccus* sp. strain R114 was subjected to mutagenesis with ultraviolet (UV) light followed by screening for red colonies. The UV mutagenesis was performed as follows.

10 An overnight culture of strain R114 was grown in ME medium (see Example 2). The overnight culture was subcultured into fresh ME medium (initial OD₆₁₀ = 0.1) and incubated at 28°C for 3 hours. Aliquots of this culture were centrifuged and the pellet washed with 20mM potassium phosphate buffer (pH 7.2). After a second centrifugation, the pellet was resuspended to a final OD₆₁₀ of 0.1. Ten milliliter aliquots of the cell

15 suspension were placed in a sterile 100-ml glass beaker. The thin layer of cell suspension was irradiated with UV light at a flux of 1450μW/cm² for a pre-determined optimal length of time. The cell suspension was mixed during the irradiation by means of a paper clip in the beaker and a magnetic stirrer. The mutagenized cell suspensions (and the unmutagenized controls) were plated on 362/F2 agar medium (Table 28). Triplicate viable

20 plate counts (in dim room light) were done on suspensions before and after mutagenesis. Plates were incubated for 4-5 days at 28°C, and the colonies were scored. Several red colonies (putative lycopene producers) were identified and purified by re-streaking. One mutant, designated UV7-1, was further evaluated for lycopene production.

25 Table 29 shows the zeaxanthin production and lycopene production by the control strain R114 and its mutant derivative UV7-1. Strain R114 produced only zeaxanthin. Mutant UV7-1 produced mostly lycopene, but also produced a residual amount of zeaxanthin, suggesting that the mutational block in UV7-1 (presumably in the crtY gene) is not complete. These results show that it is possible to derive lycopene producing strains from *Paracoccus* sp. strain R114.

- 100 -

Table 28. Recipe and preparation for medium 362F/2

Component	Amount
Glucose monohydrate	33 g
Yeast extract	10 g
Tryptone	10 g
NaCl	5 g
MgSO ₄ ·7H ₂ O	2.5 g
Agar (for solid medium)	20 g
Distilled water	To 932 ml
-adjust pH to 7.4	
-sterilize by filtration (liquid medium) or autoclaving (solid medium)	
-Add 2.5 ml each of microelements solution, NKP solution and CaFe solution	
Microelements solution	Amount per liter distilled water
(NH ₄) ₂ Fe(SO ₄) ₂ ·6H ₂ O	80 g
ZnSO ₄ ·7H ₂ O	6 g
MnSO ₄ ·H ₂ O	2 g
NiSO ₄ ·6H ₂ O	0.2 g
EDTA	6 g
-sterilize by filtration	
NKP solution	Amount per liter distilled water
K ₂ HPO ₄	250 g
(NH ₄) ₂ HPO ₄	300 g
-sterilize by filtration	
CaFe solution	Amount per liter distilled water
CaCl ₂ ·2H ₂ O	75 g
FeCl ₃ ·6H ₂ O	5 g
Concentrated HCl	3.75 ml
-sterilize by filtration	

- 101 -

Table 29. Zeaxanthin and lycopene production by *Paracoccus* sp. strain R114 and its red mutant derivative UV7-1.

	Zeaxanthin (mg/l)	Lycopene (mg/l)
24 hours		
R114	36.65	0
UV7-1	3.85	20.85
48 hours		
R114	72.95	0
UV7-1	5.75	85.95
72 hours		
R114	83.9	0
UV7-1	5.85	124.55

Example 12: Model for the Industrial Production of Astaxanthin by Fermentation Using Strains Derived from *Paracoccus* sp. strain R114

Astaxanthin is a commercially important carotenoid used primarily in the aquaculture industry. EP 872,554 showed that astaxanthin production could be achieved in *E. coli* by introducing plasmids containing combinations of the cloned carotenoid (*crt*) genes from *Paracoccus* sp. strain R1534 and *Paracoccus carotinifaciens* E-396^T. Together, the cloned *crt* genes (*crtEBIYZ*) and *crtW* (β -carotene β -4 oxygenase) encoded a total biosynthetic pathway from FPP through zeaxanthin to astaxanthin. The sequences of the *P. carotinifaciens* E-396 *crtW*, *Paracoccus* sp. R1534 *crtZ*, and *Paracoccus* sp. R1534 *crtE* genes and encoded polypeptides are set forth in (SEQ ID NOs:180 and 181 (*crtW*); 182 and 184 (*crtZ*); and 184 and 185 (*crtE*)). However, it was not shown that astaxanthin production could be achieved in the *Paracoccus* sp. strain R114 host family. To demonstrate the utility of recombinant strains derived from strain R114 for astaxanthin production, the cloned *crtW* gene (SEQ ID NO:180) was introduced into strain R114 as follows.

Table 30. PCR primers used for the work described in Example 12.

Primer name	Sequence
CrtW-Nde	5' AAGGCCTCATATGAGCGCACATGCCCTGCC 3' (SEQ ID NO:186)
CrtW-Bam	5' CGGGATCCTCATGCGGTGTCCCCCTTGG 3' (SEQ ID NO:187)
CrtZ-Nde	5' AAGGCCTCATATGAGCACTTGGGCCGCAAT 3' (SEQ ID NO:188)
CrtZ-Bam	5' AGGATCCTCATGTATTGCGATCCGCCCTT 3' (SEQ ID NO:189)

- 102 -

The *crtW* gene was amplified by PCR from the cloned *crt* cluster of *Paracoccus carotinifaciens* strain E-396^T (Tsubokura et al., supra; EP 872,554) using the primers crtW-Nde and crtW-Bam (Table 30). The primers were designed such that the ATG start codon constitutes the second half of an *Nde*I site (cleavage recognition site CATATG), and a *Bam*HI site (GGATCC) was introduced immediately after the stop codon. The PCR product was cloned in the pCR®2.1-TOPO vector, resulting in plasmid TOPO-*crtW*. The *crtW* gene was excised with *Nde*I and *Bam*HI and subcloned in the *Nde*I-*Bam*HI cut vector pBBR-K-P*crtE* (described in Example 6) to create plasmid pBBR-K-P*crtE-crtW*.

Plasmid pBBR-K-P*crtE-crtW* was transferred to *Paracoccus* sp. strain R114 using a standard bacterial conjugation procedure {*E. coli* strain S17 [Priefer et al., J. Bacteriol. 163:324-330 (1985)] was the donor organism}. Transconjugants were selected on medium 362F/2 agar (Table 28) containing 50 mg/l kanamycin and purified by restreaking on the same medium. The presence of plasmid pBBR-K-P*crtE-crtW* in the strain was confirmed by PCR. Carotenoid production by strains R114 (host control), R114/pBBR-K (empty vector control) and R114/ pBBR-K-P*crtE-crtW* was measured in shake flask cultures as described in Examples 1 and 2, except that liquid 362F/2 medium was used instead of ME medium. These results are shown in Table 31. The control strains R114 and R114/pBBR-K produced only zeaxanthin. In strain R114/ pBBR-K-P*crtE-crtW*, the zeaxanthin was completely consumed by the plasmid-encoded β-carotene β-4 oxygenase. However, although astaxanthin was produced, two other ketocarotenoids, adonixanthin and canthaxanthin, accumulated at higher levels. This indicated an imbalance *in vivo* of the β-carotene hydroxylase (encoded by the chromosomal *crtZ* gene in strain R114) and the cloned β carotene β-4 oxygenase (CrtW).

To test this hypothesis, two new plasmids were created that contained the *crtZ* and *crtW* genes together in mini-operons. The order of the genes was made different in the two constructs (i.e., *crtZ-crtW* and *crtW-crtZ*) to try and create different ratios of expression of the *crtZ* and *crtW* genes. The construction of the new plasmids required the assembly of a special set of cloning vectors as follows. A series of operon construction vectors (based on the vector pCR®2.1-TOPO) was designed to facilitate the assembly of genes (in this case, *crtZ* and *crtW*) into operons. The genes of interest must have an ATG start codon, embedded in an *Nde*I site (CATATG), and a TGA stop codon immediately followed by a *Bam*HI site.

- 103 -

Table 31. Astaxanthin production in *Paracoccus* sp. strain R114 containing plasmids expressing the *crtW* gene alone and in combination with the *crtZ* gene.

Strain	24 hours						Sp. Form. ^a
	ZXN	ADN	CXN	AXN	Total		
R114	46.5	0	0	0	46.5		2.1
R114/pBBR-K	38.8	0	0	0	41.4		2.2
R114/pBBR-K-P <i>crtE</i> - <i>crtW</i>	0	13.0	21.8	2.3	37.5		2.1
R114/pBBR-K-P <i>crtE</i> - <i>crtWZ</i>	0	14.9	29.5	1.3	45.6		2.1
R114/pBBR-K-P <i>crtE</i> - <i>crtZW</i>	0	18.0	20.4	7.3	45.65		2.1
	48 hours						
R114	72.6	0	0	0	74.4		2.8
R114/pBBR-K	70.1	0	0	0	70.1		3.1
R114/pBBR-K-P <i>crtE</i> - <i>crtW</i>	0	26.7	22.0	26.9	75.5		3.9
R114/pBBR-K-P <i>crtE</i> - <i>crtWZ</i>	0	30.9	27.2	34.8	92.9		4.0
R114/pBBR-K-P <i>crtE</i> - <i>crtZW</i>	0	15.7	11.2	58.3	85.1		4.1
	72 hours						
R114	82.5	0	0	0	82.5		5.3
R114/pBBR-K	82.9	0	0	0	82.9		5.1
R114/pBBR-K-P <i>crtE</i> - <i>crtW</i>	0	19.7	17.0	46.8	83.5		5.2
R114/pBBR-K-P <i>crtE</i> - <i>crtWZ</i>	0	28.7	26.4	43.8	98.8		6.1
R114/pBBR-K-P <i>crtE</i> - <i>crtZW</i>	0	18.3	14.4	66.3	98.9		5.9

^aZXN, zeaxanthin; AND, adonixanthin; CXN, canthaxanthin; AXN, astaxanthin.

^bSpecific Formation, expressed as mg/l total carotenoid/OD₆₆₀.

- 5 Furthermore, the first nucleotide after the start codon and the last nucleotide before the stop codon must be adenine and the gene must lack sites for at least one of the enzymes *BsgI*, *BseMII*, *BseRI* and *GsuI*. Four operon construction vectors were constructed, differing in the arrangements of their polylinker sequences (SEQ ID NOs: 190-197). The cleavage sites of the first two enzymes are within the *NdeI* site. The cleavage sites of the
- 10 last two enzymes are before the *BamHI* site. The *BseRI* site in pOCV-1 and pOCV-4 is not unique and cannot be used for operon construction.

The genes to be assembled in operons are first inserted individually between the *NdeI* and the *BamHI* sites of the appropriate operon construction vectors. The resulting plasmid with the upstream gene of the envisioned operon is then cut with one of the two enzymes

- 104 -

at the end of the polylinker and with an enzyme, which has a unique site within the vector backbone. The plasmid containing the downstream gene of the envisioned operon is cut with one of the first two enzymes of the polylinker and with the same enzyme (with a unique site in the vector backbone) used for the first plasmid (containing the desired upstream gene). The fragments carrying the genes are isolated and ligated, resulting in a pOCV plasmid with both genes between the *NdeI* and the *BamHI* sites. More genes can be added in an analogous fashion. The assembled genes overlap such that the first two nucleotides, TG, of the TGA stop codon of the upstream gene coincide the last two nucleotides of the ATG start codon of the downstream gene. The same overlap is found between all genes in the carotenoid (*crt*) operon (*crtZYIB*) in *Paracoccus* sp. strain R1534 (Pasamontes et al., supra).

The pOCV backbone is derived from pCR®2.1-TOPO. The *BseMII* site in the region necessary for replication, upstream of the *ColE1* origin, was eliminated by site directed mutagenesis changing the site from CTCAG into CACAG. The remaining three *BseMII* sites and one *GsuI* site were eliminated by removing a 0.8 kb *DdeI*-*Asp700* fragment. The remaining vector was blunt-end ligated after fill-in of the *DdeI* recessed end. The poly-linkers were inserted between the *BamHI* and *XbaI* sites by means of annealed oligo-nucleotides with the appropriate 5' overhangs.

Plasmid pBBR-K-P*crtE-crtZW*, was constructed using the operon construction vector pOCV-2 as follows. The *crtZ* gene was amplified by PCR from *Paracoccus* sp. strain R114 using the primers *crtZ-Nde* and *crtZ-Bam* (Table 30). The primers were designed such that the ATG start codon constitutes the second half of a *NdeI* site (cleavage recognition site CATATG) and a *BamHI* site (GGATCC) was introduced immediately after the stop codon. The PCR product was cloned in the pCR®2.1-TOPO vector, resulting in plasmid TOPO-*crtZ*. To assemble the two genes in a mini-operon, both genes, *crtZ* and *crtW* were excised with *NdeI* and *BamHI* from the plasmids TOPO-*crtZ* and TOPO-*crtW* and sub-cloned in the *NdeI*-*BamHI* cut vector pOCV-2, creating plasmids pOCV-2-*crtZ* and pOCV-2-*crtW*. Plasmid pOCV-2-*crtZ* was cut with *BseMII* and *PstI* (there is a unique *PstI* site in the kanamycin resistance gene) and the 2.4 kb fragment (containing *crtZ*) was ligated with the *crtW*-containing 1876 bp *BseRI*-*PstI* fragment from pOCV-2-*crtW*. The resulting plasmid, pOCV-2-*crtZW*, was cut with *NdeI* and *BamHI* and the *crtZW* fragment was ligated with the *NdeI*-*BamHI* backbone of pBBR-K-P*crtE* to yield pBBR-K-P*crtE-crtZW*. Plasmid pBBR-K-P*crtE-crtWZ*, was constructed in an analogous fashion.

The data in Table 31 show that the ratio of adonixanthin, canthaxanthin and astaxanthin did not change appreciably in strain R114/ pBBR-K-P*crtE-crtWZ* compared to strain

- 105 -

pBBR-K-P_{crtE}-c_{rW}. However, in strain pBBR-K-P_{crtE}-c_{rZW}, the production of the ketocarotenoids was shifted in favor of astaxanthin. This result indicates that the level of expression is dependent on the position of the gene within the mini-operon, and suggests that increasing the *in vivo* level of β-carotene hydroxylase activity creates a balance 5 between the activities of this enzyme and β-carotene β-4 oxygenase that is more favorable for full conversion of zeaxanthin to astaxanthin.

The results described in this Example also show that it is possible, through appropriate genetic engineering, to produce not only astaxanthin, but also other ketocarotenoids of commercial interest in *Paracoccus* sp. strain R114 or its relatives. For example, expression 10 of a gene coding for β-carotene β-4 oxygenase in a *crtZ* mutant of strain R114 (lacking β-carotene hydroxylase activity) would provide for production of exclusively ketocarotenoids, e.g., echinenone or canthaxanthin, without co-production of hydroxylated carotenoids. Taken together, the results presented in this Example and Example 11 show the broad utility of *Paracoccus* sp. strain R114 and its relatives to produce industrially im- 15 portant carotenoids.

Example 13: Accumulation of mevalonate in cultures of *Paracoccus* sp. strain R114 overexpressing genes of the mevalonate pathway

Overexpression of the genes of the mevalonate pathway in *Paracoccus* sp. strain R114 leads to increased carbon flow to through the mevalonate pathway. The construction of plasmid 20 pBBR-K-mev-op16-2 was described in Example 5. Plasmid pBBR-K-mev-op-up-4 was constructed as follows. A DNA fragment containing containing most of the *mvaA* gene and the entire *idi* and *hcs* genes was obtained on a 3.1 kb *Sma*I-*Sal*I fragment following partial digestion of a λ-clone containing the *Paracoccus* sp. strain R114 mevalonate operon (see Example 4). This fragment was subcloned in pUC19, yielding the plasmid 25 pUC19mev-op-up'. To facilitate subcloning, the *Kpn*I-*Hind*III fragment of pUC19mev-op-up' containing the mevalonate genes was recloned in the vector pBluescriptKS⁺, resulting in plasmid pBluKSp-mev-op-up'. A 1.7 kb *Sal*I fragment from pUC19mev-op-up' was then cloned in the *Sal*I site of plasmid 2ES2-1, which is a pUC19-derived plasmid containing the cloned *Sal*I-*Eco*RI fragment M from *Paracoccus* sp. strain R114 (refer to 30 Example 4). This resulted in plasmid pUC19mev-op-up-2. Plasmid pUCmev-op-up-3 was then obtained by combining the *Bbs*I-*Bsa*I fragment from pUC19mev-op-up-2 carrying the beginning of the mevalonate operon with the *Bbs*I-*Bsa*I fragment from pBluKSp-mev-op-up' containing *idi* and *hcs*. Separately, a unique *Mlu*I site was introduced between the *Nsi*I and *Kpn*I sites of the vector pBBR1MCS-2 (refer to Example 35 5) by inserting an annealed primer containing an *Mlu*I restriction site. The resulting new

- 106 -

cloning vector pBBR-K-Mlu was cut with *Mlu*I and *Kpn*I and the *Mlu*I-*Kpn*I fragment from pUCmev-op-up-3, containing the first three genes of the mevalonate operon, was inserted, yielding plasmid pBBR-K-mev-op-up-3. Plasmid pBBR-K-mev-op-up-4 was then constructed by insertion of the *Sma*I fragment from plasmid 16SB3, which contains 5 most of the *mvk* gene and the 5' end of *pmk* (plasmid 16SB3 is a pUC19-derived plasmid containing the *Paracoccus* sp. strain R114 *Sall*-*Bam*HI fragment A; refer to Example 4). The insert of plasmid pBBR-K-mev-op-up-4 contains the putative mevalonate operon promoter region, the first four genes of the mevalonate operon and the 5' end of *pmk*.

Plasmids pBBR-K-mev-op16-2 and pBBR-K-mev-op-up-4 were each introduced into 10 *Paracoccus* sp. strain R114 by electroporation. Production of zeaxanthin and mevalonate by the new strains were compared to the control strain R114. The strains were grown in baffled shake flasks in liquid medium 362F/2 (see Example 11) for 72 hours. For strains R114/pBBR-K-mev-op16-2 and R114/pBBR-K-mev-op-up-4, kanamycin (50 mg/l) was also added to the cultures. The cultivation temperature was 28°C and shaking was at 200 15 rpm. Zeaxanthin was measured by the method set forth in Example 1, while mevalonate in the culture supernatants was measured as follows: A 0.6 ml sample of the culture was centrifuged for 4 minutes at 13,000 x g. Four hundred microliters of the supernatant were added to 400 microliters of methanol and mixed by vortexing for 1 min. The mixture was centrifuged again for 4 minutes at 13,000 x g. The resulting supernatant was then analyzed 20 directly by gas chromatography (GC) using the method of Lindemann et al. [J. Pharm. Biomed. Anal. 9:311-316 (1991)] with minor modification as follows. The GC was a Hewlett-Packard 6890plus instrument (Hewlett-Packard, Avondale, PA, USA) equipped with a cool-on-column injector and a flame ionization detector. One microliter of sample prepared as described above was injected onto a fused silica capillary column (15m length 25 x 0.32mm ID) coated with a 0.52 micron film of crosslinked modified polyethylene glycol (HP-FFAP, Agilent Technologies, USA). Helium was used as the carrier gas at an inlet pressure of 0.6 bar. The temperature of the programmable injector was ramped from 82°C to 250°C at a rate of 30°C/minute. The column temperature profile was 80°C for 0.5 minutes, followed by a linear temperature gradient at 15°C/min to 250°C and finally held 30 at 250°C for 5 minutes. The detector temperature was maintained at 320°C.

In the first experiment, zeaxanthin and mevalonate production were measured in strains R114 and R114/pBBR-K-mev-op16-2 (Table 32). Both strains produced similar amounts of zeaxanthin, but strain R114/pBBR-K-mev-op16-2 produced a four-fold higher level of mevalonate. These results show that overexpression of the genes of the mevalonate pathway in *Paracoccus* sp. strain R114 results in increased carbon flow through the mevalonate 35

- 107 -

pathway. The accumulation of mevalonate was expected because strain R114/pBBR-K-mev-op16-2 does not have an overexpressed *crtE* gene, and the *crtE* gene product (GGPP synthase) is known to be a limiting step in zeaxanthin production in *Paracoccus* sp. strain R114 (see Examples 6 and 8). Cells having a limiting amount of GGPP synthase, upon 5 overproduction of the enzymes of the mevalonate pathway, would be expected to accumulate FPP, and it is well known that FPP is a potent inhibitor of mevalonate kinase [Dorsey and Porter, J. Biol. Chem. 243:4667-4670 (1968); Gray and Kekwick, BBA 279:290-296 (1972); Hinson et al. J. Lipids Res. 38:2216-2223 (1997)]. Therefore, accumulation of 10 FPP resulting from overexpression of the genes of the mevalonate pathway would cause inhibition of mevalonate kinase, which in turn is manifested as mevalonate accumulation in the culture.

Table 32. Zeaxanthin and mevalonate production in strains R114 and R114/pBBR-K-mev-op16-2.

Strain/plasmid	Mevalonate (mg/l)	Zeaxanthin(mg/l)
R114	50.5	70.0
R114/pBBR-K-mev-op16-2	208.2	65.2

15 In a second experiment, zeaxanthin and mevalonate production were measured in strain R114 and two independent isolates of R114/pBBR-K-mev-op-up-4 (Table 33). These results again show that overexpression of the genes of the mevalonate pathway increased carbon flow through the mevalonate pathway.

Table 33. Zeaxanthin and mevalonate production in strains R114 and R114/pBBR-K-mev-op-up-4.

Strain/plasmid	Mevalonate (mg/l)	Zeaxanthin(mg/l)
R114	45	67.5
R114/pBBR-K-mev-op-up-4 (Isolate 1)	133.2	53.7
R114/pBBR-K-mev-op-up-4 (Isolate 2)	163.7	47.6

- 108 -

The following biological material was deposited under the terms of the Budapest Treaty with the American Type Culture Collection (ATCC) at 10801 University Blvd., Manassas, VA 20110-2201, USA, and were assigned the following accession numbers:

Strain	Accession No.	Date of Deposit
<i>Paracoccus</i> sp. R114	PTA-3335	April 24, 2001
<i>Paracoccus</i> sp. R1534	PTA-3336	April 24, 2001
<i>Paracoccus</i> sp. R-1506	PTA-3431	June 5, 2001

- 5 All patents, patent applications, and publications cited above are incorporated herein by reference in their entirety as if recited in full herein.

The invention being thus described, it will be obvious that the same may be varied in many ways. Such variations are not to be regarded as a departure from the spirit and scope of the invention and all such modifications are intended to be included within the scope of
10 the following claims.

What is Claimed Is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown as residues 1 to 340 of SEQ ID NO:43;
 - 5 (b) an amino acid sequence shown as residues 1 to 349 of SEQ ID NO:45;
 - (c) an amino acid sequence shown as residues 1 to 388 of SEQ ID NO:47;
 - (d) an amino acid sequence shown as residues 1 to 378 of SEQ ID NO:49;
 - (e) an amino acid sequence shown as residues 1 to 305 of SEQ ID NO:51;
 - (f) an amino acid sequence shown as residues 1 to 332 of SEQ ID NO:53;
 - 10 (g) a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOs: 43, 45, 47, 49, 51, and 53, wherein said fragment has at least 30 contiguous amino acid residues;
 - (h) an amino acid sequence of a fragment of a polypeptide selected from the group consisting of SEQ ID NOs: 43, 45, 47, 49, 51, and 53, the fragment having the activity of
 - 15 hydroxymethylglutaryl-CoA reductase (HMG-CoA reductase), isopentenyl diphosphate isomerase, hydroxymethylglutaryl-CoA synthase (HMG-CoA synthase), mevalonate kinase, phosphomevalonate kinase, or diphosphomevalonate decarboxylase;
 - (i) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe comprising at least 30 consecutive
 - 20 nucleotides of SEQ ID NO:42 or a complement of SEQ ID NO:42, wherein the polypeptide has the activity of HMG-CoA reductase, isopentenyl diphosphate isomerase, HMG-CoA synthase, mevalonate kinase, phosphomevalonate kinase, or diphosphomevalonate decarboxylase; and
 - (j) a conservatively modified variant of SEQ ID NO:43, 45, 47, 49, 51 or 53.
- 25 2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown as residues 1 to 287 of SEQ ID NO:159;
 - (b) at least 30 contiguous amino acid residues of SEQ ID NO:159;
 - (c) an amino acid sequence of a fragment of SEQ ID NO: 159, the fragment having the
 - 30 activity of farnesyl diphosphate synthase (FPP synthase);
 - (d) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe comprising at least 30 consecutive nucleotides spanning positions 295-1158 of SEQ ID NO:157 or a complement thereof, wherein the polypeptide has the activity of FPP synthase; and
 - (e) a conservatively modified variant of SEQ ID NO:159.

3. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown as residues 1 to 142 of SEQ ID NO:160;
 - (b) at least 30 contiguous amino acid residues of SEQ ID NO:160;
 - 5 (c) an amino acid sequence of a fragment of SEQ ID NO: 160, the fragment having the activity of 1-deoxyxylulose-5-phosphate synthase (DXPS);
 - (d) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe comprising at least 30 consecutive nucleotides spanning positions 1185-1610 of SEQ ID NO:157 or a complement thereof,
 - 10 wherein the polypeptide has the activity of DXPS;
 - (e) a conservatively modified variant of SEQ ID NO:160.
4. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown as residues 1 to 390 of SEQ ID NO:178;
 - 15 (b) at least 30 contiguous amino acid residues of SEQ ID NO:178;
 - (c) an amino acid sequence of a fragment of a polypeptide of SEQ ID NO: 178, the fragment having the activity of acetyl-CoA acetyltransferase;
 - (d) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes under stringent conditions to a hybridization probe comprising at least 30 consecutive
 - 20 nucleotides spanning positions 1-1170 of SEQ ID NO:177 or a complement thereof, wherein the polypeptide has the activity of acetyl-CoA acetyltransferase; and
 - (e) a conservatively modified variant of SEQ ID NO:178.
5. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown as residues 1 to 240 of SEQ ID NO:179;
 - 25 (b) at least 30 contiguous amino acid residues of SEQ ID NO:179;
 - (c) an amino acid sequence of a fragment of a polypeptide of SEQ ID NO: 179, the fragment having the activity of acetoacetyl-CoA reductase;
 - (d) an amino acid sequence of a polypeptide encoded by a polynucleotide that hybridizes
 - 30 under stringent conditions to a hybridization probe comprising at least 30 consecutive nucleotides spanning positions 1258-1980 of SEQ ID NO:177 or a complement thereof, wherein the polypeptide has the activity of acetoacetyl-CoA reductase; and
 - (e) a conservatively modified variant of SEQ ID NO:179.
6. An isolated polynucleotide sequence comprising a nucleotide sequence selected from the
- 35 group consisting of SEQ ID NO:42, variants of SEQ ID NO:42 containing one or more

- substitutions according to the *Paracoccus* sp. strain R1534 codon usage table, fragments of SEQ ID NO:42 that encode a polypeptide having an activity selected from the group consisting of hydroxymethylglutaryl-CoA reductase (HMG-CoA reductase), isopentenyl diphosphate isomerase, hydroxymethylglutaryl-CoA synthase (HMG-CoA synthase),
- 5 mevalonate kinase, phosphomevalonate kinase, and diphosphomevalonate decarboxylase, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides of SEQ ID NO:42, or the complement of SEQ ID NO:42, which polynucleotide encodes a polypeptide having an activity selected from the group consisting of HMG-CoA reductase,
- 10 isopentenyl diphosphate isomerase, HMG-CoA synthase, mevalonate kinase, phosphomevalonate kinase, and diphosphomevalonate decarboxylase.
7. An isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of the nucleotide sequence of SEQ ID NO:157, variants of SEQ ID NO:157 containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table, fragments of SEQ ID NO:157 that encode a polypeptide having farnesyl diphosphate (FPP) synthase activity, 1-deoxy-D-xylulose 5-phosphate synthase activity or a polypeptide having the activity of XseB, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides of SEQ ID NO:157, or the complement of SEQ ID NO:157, which polynucleotide encodes a polypeptide having an activity selected from the group consisting of FPP synthase activity, 1-deoxy-D-xylulose 5-phosphate synthase activity, and the activity of XseB.
8. An isolated polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of the nucleotide sequence of SEQ ID NO:177, variants of SEQ ID NO:177 containing one or more substitutions according to the *Paracoccus* sp. strain R1534 codon usage table, fragments of SEQ ID NO:177 that encode a polypeptide having an activity selected from the group consisting of acetyl-CoA acetyltransferase and acetoacetyl-CoA reductase, and polynucleotide sequences that hybridize under stringent conditions to a hybridization probe the nucleotide sequence of which consists of at least 30 contiguous nucleotides of SEQ ID NO:177, or the complement of SEQ ID NO:177, which polynucleotide encodes a polypeptide having an activity selected from the group consisting of acetyl-CoA acetyltransferase and acetoacetyl-CoA reductase.
9. An isolated polynucleotide sequence comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:42, SEQ ID NO:157, SEQ ID NO:177, and combinations thereof.

10. An expression vector comprising the polynucleotide sequence according to claim 6, 7, 8 or 9.
11. An expression vector selected from the group consisting of pBBR-K-mev-op16-1, pBBR-K-mev-op16-2, pDS-*mvaA*, pDS-*idi*, pDS-*hcs*, pDS-*mvk*, pDS-*pmk*, pDS-*mvd*, pDS-*His-mvaA*, pDS-*His-idi*, pDS-*His-hcs*, pDS-*His-mvk*, pDS-*His-pmk*, pDS-*His-mvd*, pBBR-K-Zea4, pBBR-K-Zea4-up, pBBR-K-Zea4-down, pBBR-K-PcrtE-crtE-3, pBBR-tK-PcrtE-*mvaA*, pBBR-tK-PcrtE-*idi*, pBBR-tK-PcrtE-*hcs*, pBBR-tK-PcrtE-*mvk*, pBBR-tK-PcrtE-*pmk*, pBBR-tK-PcrtE-*mvd*, pBBR-K-PcrtE-*mvaA-crtE-3*, pDS-*His-phaA*, pBBR-K-PcrtE-*crtW*, pBBR-K-PcrtE-*crtWZ*, pBBR-K-PcrtE-*crtZW*, and combinations thereof.
- 10 12. A cultured cell comprising the polynucleotide sequence according to claim 6, 7, 8 or 9, or an expression vector according to claim 10 or 11, or a progeny of the cell, wherein the cell expresses a polypeptide encoded by the polynucleotide sequence.
13. A method of producing a carotenoid comprising culturing a cell according to claim 12 under conditions permitting expression of a polypeptide encoded by the polynucleotide sequence, and isolating the carotenoid from the cell or the medium of the cell.
14. A method of making a carotenoid-producing cell comprising:
 - (a) introducing into a cell a polynucleotide sequence encoding an enzyme in the mevalonate pathway, which enzyme is expressed in the cell; and
 - (b) selecting a cell containing the polynucleotide sequence of step (a) that produces a carotenoid at a level that is about 1.1-1,000 times the level of the carotenoid produced by the cell before introduction of the polynucleotide sequence.
15. A method for engineering a bacterium to produce an isoprenoid compound comprising:
 - (a) culturing a parent bacterium in a medium under conditions permitting expression of an isoprenoid compound, and selecting a mutant bacterium from the culture medium that produces about 1.1-1,000 times more of an isoprenoid compound than the parent bacterium;
 - (b) introducing into the mutant bacterium an expression vector comprising a polynucleotide sequence represented by SEQ ID NO:42 operably linked to an expression control sequence; and
 - (c) selecting a bacterium that contains the expression vector and produces at least about 1.1 times more of an isoprenoid compound than the mutant in step (a).

- 113 -

16. A microorganism of the genus *Paracoccus*, which microorganism has the following characteristics:
- (i) a sequence similarity to SEQ ID NO:12 of >97% using a similarity matrix obtained from a homology calculation using GeneCompar v. 2.0 software with a gap penalty of 0%;
- 5 a homology to strain R-1512, R1534, R114 or R-1506 of >70% using DNA:DNA hybridization at 81.5°C;
- a G+C content of its genomic DNA that varies less than 1% from the G+C content of the genomic DNA of R114, R-1512, R1534, and R-1506; and
- 10 an average DNA fingerprint that clusters at about 58% similarity to strains R-1512, R1534, R114 and R-1506 using the AFLP procedure of Example 2, with the proviso that the microorganism is not *Paracoccus* sp. (MBIC3966);
- (ii) 18:1w7c comprising at least about 75% of the total fatty acids of the cell membranes; an inability to use adonitol, i-erythritol, gentiobiose, β-methylglucoside, D-sorbitol, xylitol and quinic acid as carbon sources for growth; and
- 15 an ability to use L-asparagine and L-aspartic acid as carbon sources for growth, with the proviso that the microorganism is not *Paracoccus* sp. (MBIC3966); or
- (iii) an ability to grow at 40°C; an ability to grow in a medium having 8% NaCl; an ability to grow in a medium having a pH of 9.1; and
- 20 a yellow-orange colony pigmentation, with the proviso that the microorganism is not *Paracoccus* sp. (MBIC3966).

SEQUENCE LISTING

<110> BERRY, Alan

BRETZEL, Werner

HUMBELIN, Markus

LOPEZ-ULIBARRI, Rual

MAYER, Anne

YELISEEV, Alexei

<120> IMPROVED ISOPRENOID PRODUCTION

<130> C38435/121966

<160> 197

<170> PatentIn version 3.0

<210> 1

<211> 20

<212> DNA

<213> synthetic construct

<400> 1

agagtttgat cctggctcag

20

<210> 2

<211> 20

<212> DNA

<213> synthetic construct

<220>

<221> misc_feature

<223> n is c or t

<400> 2

ctggctcagg angaacgctg

20

<210> 3

<211> 20

<212> DNA

<213> synthetic construct

<400> 3

aaggaggtga tccagccgca

20

<210> 4

<211> 20

<212> DNA

<213> synthetic construct

<400> 4

ctcctacggg aggcagcagt

20

<210> 5

<211> 18

<212> DNA

<213> synthetic construct

<400> 5

cagcagccgc ggtaatac

18

<210> 6

<211> 19

<212> DNA

<213> synthetic construct

<400> 6

aactcaaagg aattgacgg

19

<210> 7

<211> 20

<212> DNA

<213> synthetic construct

<400> 7

agtcccgcaa cgagcgcaac

20

<210> 8

<211> 20

<212> DNA

<213> synthetic construct

<400> 8

gctacacacg tgctacaatg

20

<210> 9

<211> 20

<212> DNA

<213> synthetic construct

<400> 9

actgctgcct cccgttaggag

20

<210> 10

<211> 18

<212> DNA

<213> synthetic construct

<400> 10

gttattaccgc ggctgctg

18

<210> 11

<211> 20

<212> DNA

<213> synthetic construct

<400> 11

gttgcgcgtcg ttgcgggact

20

<210> 12

<211> 1404

<212> DNA

<213> Paracoccus sp. R-1512

<400> 12

gcggcaggct taacacatgc aagtgcagcg aggtcttcgg acctagcggc ggacgggtga 60

gtaacgcgtg ggaacgtgcc ctttgctacg gaatagtccc gggaaaactgg gttaataacc 120

gtatgtgccc tacggggaa agatttatcg gcaaaggatc ggcccgcggtt ggatttaggt 180

gttggtgtggg taatggccta ccaagccgac gatccatagc tggtttgaga ggatgatcag 240

ccacactggg actgagacac ggcccagact cctacgggag gcagcagtgg ggaatcttag 300

acaatggggg caaccctgat ctggccatgc cgctgtgatg atgaaggccc tagggttgta 360

aagctcttcc agctggaaag ataatgacgg taccagcaga agaagccccg gctaactccg 420

tgccagcagc cgccgtataa cggagggggc tagcgttggtt cggaattact gggcgtaaag 480

cgcacgtagg cggactggaa agttgggggt gaaatcccgg ggctcaacct cggaactgcc 540

tccaaaacta tcagtctgga gttcgagaga ggtgagtgaa ataccgagtg tagaggtgaa 600

attcgtagat attcggtgga acaccagtgg cgaaggcggc tcactggctc gatactgacg 660

ctgagggtgcg aaagcgtggg gagcaaacag gattagatac cctggtagtc cacgccgtaa	720
acgatgaatg ccagtcgtcg ggttgcattgc aattcggtga cacacctaac ggattaagca	780
ttccgcctgg ggagtacggt cgcaagatta aaactcaaag gaattgacgg gggcccccac	840
aagcgggtgga gcatgtggtt taattcgaag caacgcgcag aaccttacca acccttgaca	900
tccctggaca tcccggagaga tcgggcttgc acttcggtga ccaggagaca ggtgctgcat	960
ggctgtcgtc agctcggtgc gtgagatgtt cggttaagtc cgccaacgag cgcaaccac	1020
gtcccttagtt gccagcattc agttgggcac tctatggaaa ctgccgatga taagtccggag	1080
gaaggtgtgg atgacgtcaa gtcctcatgg cccttacggg ttgggctaca cacgtgctac	1140
aatgggtggtg acagtgggtt aatccccaaa agccatctca gttcggatttgc tccctctgcaa	1200
ctcgaggggca tgaagttgga atcgctagta atcgccgaac agcatgcccgc ggtgaatacg	1260
ttccccgggcc ttgtacacac cgcccggtcac accatgggag ttggttctac ccgacgacgc	1320
tgcgctaacc cttcggggag gcaggcggcc acggtaggat cagcgactgg ggtgaagtcg	1380
taacaaggta gccgttagggg aacc	1404

<210> 13

<211> 20

<212> DNA

<213> synthetic construct

<400> 13

tcgttagactg cgtacaggcc

20

<210> 14

<211> 14

<212> DNA

<213> synthetic construct

<400> 14

catctgacgc atgt

14

<210> 15

<211> 16

<212> DNA

<213> synthetic construct

<400> 15

gacgatgagt cctgac

16

<210> 16

<211> 14

<212> DNA

<213> synthetic construct

<400> 16

tactcaggac tggc

14

<210> 17

<211> 17

<212> DNA

<213> synthetic construct

<400> 17

gactgcgtac aggccca

17

<210> 18

<211> 18

<212> DNA

<213> synthetic construct

<400> 18

cgatgagtcc tgaccgaa

18

<210> 19

<211> 18

<212> DNA

<213> synthetic construct

<400> 19

cgatgagtcc tgaccgac

18

<210> 20

<211> 17

<212> DNA

<213> synthetic construct

<400> 20

gactgcgtac aggcccc

17

<210> 21

<211> 17

<212> DNA

<213> synthetic construct

<400> 21

gactgcgtac aggcccg

17

<210> 22

<211> 18

<212> DNA

<213> synthetic construct

<400> 22

cgatgagtcc tgaccgag

18

<210> 23

<211> 8

<212> PRT

<213> Paracoccus sp. R1534

<220>

<221> UNSURE

<222> (2)..(2)

<223> Xaa is Leu or Ile

<400> 23

Ala Xaa Ile Lys Tyr Trp Gly Lys
1 5

<210> 24

<211> 27

<212> DNA

<213> Paracoccus sp. R1534

<400> 24

ccsctgatca artaytgggg baaratc

27

<210> 25

<211> 20

<212> DNA

<213> Paracoccus sp. R1534

<400> 25

gcsctgatca artaytgggg

20

<210> 26

<211> 20

<212> DNA

<213> Paracoccus sp. R1534

<400> 26
gcsatcatca artaytgggg

20

<210> 27

<211> 17

<212> DNA

<213> Paracoccus sp. R1534

(C) <400> 27
atcaartayt gggtaaa

17

<210> 28

<211> 17

<212> DNA

<213> Paracoccus sp. R1534

<400> 28
atcaartayt ggggcaa

17

<210> 29

<211> 17

<212> DNA

<213> Paracoccus sp. R1534

<400> 29
atcaartayt gggggaa

17

<210> 30

<211> 17

<212> DNA

<213> Paracoccus sp. R1534

<400> 30
atcaartayt ggggaaa

17

<210> 31
<211> 8
<212> PRT
<213> Paracoccus sp. R1534

<220>
<221> UNSURE
<222> (7)..(7)
<223> Xaa is Asn or Gln

<400> 31
Thr Met Asp Ala Gly Pro Xaa Val
1 5

<210> 32
<211> 24
<212> DNA
<213> Paracoccus sp. R1534

<220>
<221> misc_feature
<222> (19)..(19)
<223> n is a or c

<220>
<221> misc_feature
<222> (21)..(21)

<223> n is y or r

<400> 32
acsatggayg csggbccsna ngts

24

<210> 33

<211> 24

<212> DNA

<213> Paracoccus sp. R1534

()
<220>

<221> misc_feature

<222> (19)..(19)

<223> n is t or g

<220>

<221> misc_feature

<222> (21)..(21)

<223> n is r or y

()
<400> 33
tgstacctrac gsccvuggsnt ncas

24

<210> 34

<211> 17

<212> DNA

<213> Paracoccus sp. R1534

<400> 34
tggtagcttac gsccvagg

17

<210> 35

<211> 17

<212> DNA

<213> Paracoccus sp. R1534

<400> 35

tggtaacctgc gscsvgg

17

<210> 36

<211> 17

<212> DNA

<213> Paracoccus sp. R1534

<400> 36

tgctaacctac gscsvgg

17

<210> 37

<211> 17

<212> DNA

<213> Paracoccus sp. R1534

<400> 37

tgctaacctgc gscsvgg

17

<210> 38

<211> 20

<212> DNA

<213> Paracoccus sp. R1534

<400> 38

tacctacgsc cvggsttrca

20

<210> 39

<211> 20

<212> DNA

<213> Paracoccus sp. R1534

<400> 39

tacctgcgsc cvggsttrca

20

<210> 40

<211> 20

<212> DNA

<213> Paracoccus sp. R1534

<400> 40

tacctacgsc cvggsgtca

20

<210> 41

<211> 20

<212> DNA

<213> Paracoccus sp. R1534

<400> 41

tacctgcgsc cvggsgtca

20

<210> 42

<211> 9066

<212> DNA

<213> Paracoccus sp. R114

<220>

<221> CDS

<222> (2622) .. (3644)

<223> mvaA gene

<400> 42
 ggatccggca gctcgacacg ccgcagaacc tgtacgaacg tccccccagc cgcttcgtcg 60
 cggaaattcgt cgggcgcggg acgggttgtc ccgtgcaggc ccatgacggc gcgggcccgcg 120
 cccgcatttctt gggggcccgag gtggcggtga acgcccggcc gcaatcgccg tttgtcgatc 180
 acgtctgcctt ggcggcccgag aaccttgcca tctccgagac gggcgacctg cgcgccaaagg 240
 tcgcgcgcgt cacctatctt ggccggaaat acctgctgga aaccgtgctg gattgcggca 300
 cccggctgggt gaccgagacc cgcccccgt tcgatacggg cgccgcagctt ggccctgacca 360
 tcaacgcccc ctgggccttt gccgaggatt gaatggacag cgtgaagatc ctttcggca 420
 tgggcgtgaa gggccctgcc tgcatcaggc tggatgtcgg cggatgcgc ctgatcctcg 480
 attgcgggac cggcccgac gagggcgccg agttcgaccc cgcctggctg gcggacgcgg 540
 atgcgggtgct gatcacccat gaccacgtgg accatatcgg cggcgccgt cacgcggctg 600
 cggcggggctt gccgatccat gcgacgcggc agacggcgaa gttgctgccc gcggggggcg 660
 atctgcgcctt gctgcccggaa cgccgggtca cgccggatcgc cggggtcgt ctgacgaccg 720
 gtcgcaacgg gcatgccgcg ggccggctt ggtgcattt cgacatgggc gaggggctgt 780
 tctattccgg cgactggtcc gaggaatccg actggttcgc cttcgatccg ccccccgcctg 840
 cggggacggc gattctcgac tgctcctatg gcgggttcga cgtggcgcaa tcggattgca 900
 tcgcggacctt ggacgacccgt ctcgagggtc tgccggggca ggtactgctg cccggccgc 960
 catccggcccg cgcggcccgag ctggccctgc ggctgatccg cgcggccacggc cccggcagcg 1020
 tgatggtcga cgacgcctgc ctgcccggca tcgcgcactt gcccggaggcg cgcggactgg 1080
 cctacgcccac cgaggcacgc ttcttgcgt ggcacacgcgaa gaacgcccggc agccggcg 1140
 gcatggcgcc atctgcaagc atggcgcat gcgggcaggc tggggcgaaa cgcgcattgtc 1200
 gtcttcaccg ggcacatgaa cgtccatgcg cgcgcatttgc ggcacccggc cggccggcat 1260
 ttccggccgtt ggaacgtgca tccggccgtc cgcgcatttgc gacggatgtc ggaacggctg 1320
 ggcgcgcggc gctttggccc ggccttctgc cccgacccggc agatctatct ggcgcgtggac 1380
 atgggcgcgc aggtcttcat gcaccaggag gtgcgcatttgc gatccccggc cgcgcatttgc 1440
 gcctgatccg ccacggcgaa acgaccggca atgcaggggc gatcatcgcc ggcgcacccg 1500
 atgtgcccctt gacgccaagg ggccgcgtc aggccccggc cctggcaggc cgcgaatggc 1560
 catcgggcatttgc cgcgcgttgc ggcacccggc tgcgcgttc cgcgcataacc ggcgcgtgg 1620

cctttccggg gcgcgaccac cagcccgaac ccgatctgcg cgaacgcgac tggggcatct	1680
tcgagggacg ccccgtcgcc gatctgcccc cgcgcaaat cacgcccgaag gggggcgagg	1740
gctgggacga cgtgatggcc cgcgtggacc gcgcgatccg gcggatctgc gcgacctcgg	1800
gcgatgcgct gccggtgctg gtctgccatt cggcggtgat ccgtgcccg cgcggtgtgt	1860
ggaccaccgg cgatgcgggc gatgtccgc ccaacgccac gccgatcctg ttcagcccg	1920
acggcgaccg attaaaggaa ggaacgatat gaccgccacc acccccgtcg tcgttccga	1980
acgtggacgg cacgcttgcc gaattcgacg ccgaccgcct gggccatctt gtccacggca	2040
cgaccaagca ctgggacgcc ttccaccacg cgatggccga cgccccgccc atccccgagg	2100
tcgcccgcct gatgcgcaag ctgaaggagg gggcgagac ggtcgtcatac tgctcgggc	2160
ggccccgcgg ctggcaggat cagacgatcg catggctgcg caagcacgac ctgcccattcg	2220
acgggatcta tctgcgcccc gaggatcagg acggcgccag cgaccccgag gtcaagcgcc	2280
gcgccttagc cgagatgcgc gccgacggc tggcgccctg gctggtcgtg gacgaccggc	2340
ggtcgcgtgt ggatgcctgg cggccgagg ggctggctcg cctgcaatgc ggcgcgggg	2400
acttctaggg ccgcgcgacg gggcgccga caggctggc gggaaaccgc cccgccacca	2460
tgtcctgcac gcgtcgaacc gcccgtccga cgccggtttc cgcacggaaa cgcgccgaa	2520
gttgacataa ctgcacgcg acgtctcgat tctgcccgcg aagaatgcga tgcacccaga	2580
tgatgcagaa cgaagaagcg gaagcgcccg tgaaagacca g atg att tcc cat acc Met Ile Ser His Thr	2636
1 5	
 ccg gtg ccc acg caa tgg gtc ggc ccg atc ctg ttc cgc ggc ccc gtc Pro Val Pro Thr Gln Trp Val Gly Pro Ile Leu Phe Arg Gly Pro Val	2684
10 15 20	
 gtc gag ggc ccg atc agc gcg ccg ctg gcc acc tac gag acg ccg ctc Val Glu Gly Pro Ile Ser Ala Pro Leu Ala Thr Tyr Glu Thr Pro Leu	2732
25 30 35	
 tgg ccc tcg acc gcg cgg ggg gca ggg gtt tcc cgg cat tcg ggc ggg Trp Pro Ser Thr Ala Arg Gly Ala Gly Val Ser Arg His Ser Gly Gly	2780
40 45 50	
 atc cag gtc tcg ctg gtc gac gaa cgc atg agc cgc tcg atc gcg ctg Ile Gln Val Ser Leu Val Asp Glu Arg Met Ser Arg Ser Ile Ala Leu	2828
55 60 65	
 cgg gcg cat gac ggg gcg gcg acc gcc gcc tgg cag tcg atc aag Arg Ala His Asp Gly Ala Ala Ala Thr Ala Ala Trp Gln Ser Ile Lys	2876
70 75 80 85	

gcc cgc cag gaa gag gtc gcg gcc gtg gtc gcc acc acc agc cgc ttc Ala Arg Gln Glu Glu Val Ala Ala Val Val Ala Thr Thr Ser Arg Phe 90 95 100	2924
gcc cgc ctt gtc gag ctg aat cgc cag atc gtg ggc aac ctg ctt tac Ala Arg Leu Val Glu Leu Asn Arg Gln Ile Val Gly Asn Leu Leu Tyr 105 110 115	2972
atc cgc atc gaa tgc gtg acg ggc gac gcc tcg ggt cac aac atg gtc Ile Arg Ile Glu Cys Val Thr Gly Asp Ala Ser Gly His Asn Met Val 120 125 130	3020
acc aag gcc gcc gag gcc gtg cag ggc tgg atc ctg tcg gaa tac ccg Thr Lys Ala Ala Glu Ala Val Gln Gly Trp Ile Leu Ser Glu Tyr Pro 135 140 145	3068
atg ctg gcc tat tcc acg atc tcg ggg aac ctg tgc acc gac aag aag Met Leu Ala Tyr Ser Thr Ile Ser Gly Asn Leu Cys Thr Asp Lys Lys 150 155 160 165	3116
gcg tcg gcg gtc aac ggc atc ctg ggc cgc ggc aaa tac gcc gtc gcc Ala Ser Ala Val Asn Gly Ile Leu Gly Arg Gly Lys Tyr Ala Val Ala 170 175 180	3164
gag gtc gag atc ccg cgc aag atc ctg acc cgc gtg ctg cgc acc agc Glu Val Glu Ile Pro Arg Lys Ile Leu Thr Arg Val Leu Arg Thr Ser 185 190 195	3212
gcc gag aag atg gtc cgc ctg aac tac gag aag aac tat gtc ggg ggt Ala Glu Lys Met Val Arg Leu Asn Tyr Glu Lys Asn Tyr Val Gly Gly 200 205 210	3260
acg ctg gcg ggg tcg ctg cgc agt gcg aac gcg cat ttc gcc aac atg Thr Leu Ala Gly Ser Leu Arg Ser Ala Asn Ala His Phe Ala Asn Met 215 220 225	3308
ctg ctg ggc ttc tac ctg gcg acg ggg cag gac gcg gcc aac atc atc Leu Leu Gly Phe Tyr Leu Ala Thr Gly Gln Asp Ala Ala Asn Ile Ile 230 235 240 245	3356
gag gcc agc cag ggc ttc gtc cat tgc gag gcc cgc ggc gag gat ctg Glu Ala Ser Gln Gly Phe Val His Cys Glu Ala Arg Gly Glu Asp Leu 250 255 260	3404
tat ttc tcg tgc acg ctg ccc aac ctc atc atg ggc tcg gtc ggt gcc Tyr Phe Ser Cys Thr Leu Pro Asn Leu Ile Met Gly Ser Val Gly Ala 265 270 275	3452
ggc aag ggc atc ccc tcg atc gag gag aac ctg tcg cgg atg ggc tgc Gly Lys Gly Ile Pro Ser Ile Glu Glu Asn Leu Ser Arg Met Gly Cys 280 285 290	3500
cgc cag ccg ggc gaa ccc ggc gac aac gcg cgc cgt ctt gcg gcg atc Arg Gln Pro Gly Glu Pro Gly Asp Asn Ala Arg Arg Leu Ala Ala Ile 295 300 305	3548
tgc gcg ggc gtc gtg ctg tgt ggt gaa ttg tcg ctg ctt gcg gcc cag	3596

Cys Ala Gly Val Val Leu Cys Gly Glu Leu Ser Leu Leu Ala Ala Gln			
310	315	320	325
acc aac ccc gga gag ttg gtc cgc acc cac atg gag atg gag cga tga	3644		
Thr Asn Pro Gly Glu Leu Val Arg Thr His Met Glu Met Glu Arg			
330	335	340	
ccgacagcaa ggatcaccat gtcgcgggc gcaagctgga ccatctgcgt gcattggacg	3704		
acgatgcgga tatcgaccgg ggcgacagcg gcttcgaccg catcgcgctg acccatcgcg	3764		
ccctgcccga ggtggatttc gacgccatcg acacggcgac cagttcctg ggcgtgaac	3824		
tgtccttccc gctgctgatc tcgtccatga ccggcggcac cgccgaggag atcgagcgca	3884		
tcaaccgcaa cctggccgct ggtgccgagg agggccgcgt cgccatggcg gtgggctcgc	3944		
agcgcgtgat gttcaccgac ccctcgccgc gggccagctt cgacctgcgc gccatgcgc	4004		
ccaccgtgcc gctgctggcc aatatcgccg cggtgcaactt gaacatgggg ctgggctga	4064		
aggaatgcct ggccgcgatc gaggtgctgc aggccggacgg cctgtatctg cacctgaacc	4124		
ccctgcaaga ggccgtccag cccgaggggg atcgcaactt tgccgatctg ggcagcaaga	4184		
tgcggccat cgccccgcac gttccgtgc ccgtcctgct gaaggaggtg ggctgcggcc	4244		
tgtcggccgc cgatatcgcc atcgggctgc ggcggggat ccggcatttc gacgtggccg	4304		
gtcgccggcg cacatcctgg agccggatcg agtatacgcc ccggcagcgg gccgatgacg	4364		
acctgggcct ggtttccag gactgggccc tgcagaccgt ggacgcctg cgccggccgc	4424		
ggcccgccgt tgccggccat gatggAACCA gcgtgctgat cgccagcgg ggcattccgca	4484		
acggtgtcga catggcgaaa tgcgtcatcc tggggccga catgtgcggg gtgcgcgc	4544		
ccctgctgaa agcggccaa aactcgccg aggccgttgt atccgcattc cgaaactgc	4604		
atctggagtt cggacagcc atgttccccc tggttgcgg cacgcttgc gacctgaagg	4664		
acaattcctc gcttatccgt caatgaaagt gcctaagatg accgtgacag gaatcgaagc	4724		
gatcagcttc tacacccccc agaactacgt gggactggat atccttgcg cgcatcacgg	4784		
gatcgacccc gagaagttct cgaagggat cggcaggag aaaatcgcac tgcccgcca	4844		
tgacgaggat atcgtgacca tggccgcga ggcgcgcgt ccgatcatcg aacgcgcggg	4904		
cacgcaggcc atcgacacgg ttctgttcgc caccgagagc gggatcgacc agtcgaaggc	4964		
cggccatc tatctgcgc gctgctgga cctgtcgccc aactgccgtt gcgtcgagct	5024		
gaagcaggcc tgctattccg cgacggcggc gctgcagatg gcctgcgcgc atgtccccg	5084		
caagccgac cgcaagggtgc tggtgatcgc gtccgatgtc ggcgcstatg accgcgaaag	5144		

ctcgggcgag gcgacgcagg gtgcgggcgc cgtcgccatc cttgtcagcg ccgatccaa 5204
 ggtggccgag atcggcacccg tctcgcccgt gttcacccgag gatcatcatgg atttctggcg 5264
 gccgaaccac cgccgcacgc ccctgttcga cggcaaggca tcgacgctgc gctatctgaa 5324
 cgcgcgtggc gaggcgttggc acgactatcg cgcgaatggc ggcacgagt tcgcccattt 5384
 cgcgcatttc tgctatcacg tgccgttctc gcggatggc gagaaggcga acagccacct 5444
 ggccaaggcg aacaagacgc cgggttgcgtt gggcaggtt cagacgggccc tgatctacaa 5504
 cccgcaggc gggactgct ataccgggtc gatctacccg gcattcgccct cgctgctggc 5564
 gaacgctcag gaggacctga ccggcgcgtt ggtcggtctg tttagctatg gctcggtgc 5624
 gacgggcgaa ttcttcgatg cgcggatcgc gcccggttac cgcgaccacc tgttcgcgga 5684
 acgcacatcgc gaattgctgc agatcgcac gcccgtcaca tatgacgaat acgttgccct 5744
 gtgggacgag atcgacactga cgcaggcgc gcccgcacaag ggcgcggcgtt gttcaggct 5804
 ggcaggatc gaggacgaga agcgcaccta tgcgaccgg caggcctgaa gcaggcgccc 5864
 atgcggggg caagctgatc ctgtccgggg aacattccgt gctctatggt ggcggccgc 5924
 ttgcacatggc catcgccccgc tataccgagg tgtggttcac gccgcttggc attggcgagg 5984
 ggatacgcac gacattcgcc aatctctcg cggggcgac ctattcgctg aagctgctgt 6044
 cggggttcaa gtcgcggctg gaccggcggt tcgagcagtt cctgaacggc gacctaaagg 6104
 tgcacaaggt cctgacccat cccgacgatc tggcggtcta tgcgctggcg tgcgttctgc 6164
 acgacaagcc gccggggacc gccgcgatgc cgggcacccgg cgcgatgcac cacctgcgc 6224
 gaccgggtga gctgggcagc cggacggagc tggccatccgg cgcggccatg ggtcgctcg 6284
 cggccatcgt cgcggccacc acggctctgt tcgagacgct gctggaccgg cccaagacgc 6344
 ccgaacagcg cttcgaccgc gtccgcttct gcgagcggtt gaagcacggc aaggccggc 6404
 ccatcgacgc ggcacgcgtc gtgcgcggcg ggcttgcgg cgtggccggg aacggccgg 6464
 gttcgatcag cagttcgat ttgcccggg atcacgaccc tgcgctggga cgcggctgg 6524
 actgggtact gcacggcgc cccgtcagcg ggaccggcga atgcgtcagc gcggtcgcgg 6584
 cggcgcatgg tcgcgtatgc ggcgtgtgg acgccttcgc agtctgcacc cgcgcgttgg 6644
 aggccgcgtt gctgtctggg ggcagccccg acgcccacat caccgagaac cagcgcctgc 6704
 tggAACGCA cggcgatgc cggcagcga cgcaggccct cgtggcccaag atcgaggagg 6764
 cgggtggcgc ggcacagatc tgcggcgcag gttccgtgc gggcgatcac ggcggggcgg 6824
 tcctcgatgc gattgacgcac ggcacggcga tggcttcgtt catggcgcgc catcccgacc 6884

tcgactggc gcccctgcgc atgtcgca cgggggcggc acccggcccc gcgcgcgtg	6944
cgcaaccgct gccggggcag ggctgatgga tcaggtcatc cgccgcagcg cgccgggatc	7004
ggtcatgatc acgggcgaac atgccgttgt ctatggacac cgcccatcg tcgcccggat	7064
cgagcagcgc gcccattgtga cgatcgccc gcgtgccac cgcatgtttc gcatcacctc	7124
gcagatcggg gcgcgcgcagc aggggtcgct ggacgatctg cctgcggcgc ggacctatcg	7184
cttcgtgctg gccgcacatcg cgccgacacgc gccggacctg cttgcgggt tcgacatgga	7244
catcacctcg gggatcgatc cgaggctcg ggctggatcc tcggcggcgg tgacggtcgc	7304
ctgcctcggc gcgcgtcgcc ggctggcggg gcgggggacc gaggggctgc atgacgacgc	7364
gctgcgcatac gtccgcgcca tccagggcag gggcagcggg gccgatctgg cggccagcct	7424
gcatggcggc ttctcgccct atcgccgcgc cgatggcggt gccgcgcaga tcgaggcgct	7484
tccggtgccg cggggccgt tcggcctcg ctatgcgggc tacaagaccc cgacagccga	7544
ggtgctgcgc cttgtggccg atcgatggc gggcaacgag gccgccttcg acgcgcctta	7604
ctcccgatg ggccaaagcg cagatgcgc gatccgcgc gcgcagggc tggactggc	7664
tgcattccac gacgcgtga acgaatacca gcgcctgatg gagcagctgg gcgtgtccga	7724
cgacacgctg gacgcgtaca tccgcgaggc gcgcgcacgc ggcgcgcag tcgccaagat	7784
ctccggctcg gggctgggg attgcgtgct ggcactggc gaccagccca agggttcgt	7844
gcccgcaga attgcccaga agggacttgt ttctcgatgac tgatgcgtc cgacatgaa	7904
tcgcccgtgc catggcgggc gcgaccgaca tccgagcagc cgaggcttat gcgcgcagca	7964
acatcgcgct gtcgaaatac tggggcaagc gcgcacccgc gcggAACCTT ccgctgaaca	8024
gctccgtctc gatctcggtt gcaactggg gctctcatac gcgggtcgag ggtccggca	8084
cggccacga cgaggtgcat cacaacggca cgctgctgga tccggcgcac gccttcgcgc	8144
gccgcgcgtt ggcattcgct gacctgttcc ggggggggag gcacctgccc ctgcggatca	8204
cgacgcagaa ctcgatcccg acggcggcgg ggcttgcctc gtcggcctcg ggttcgcgg	8264
cgctgacccg tgcgtggcg gggcggtcg ggctggatct ggacgacacg gatctgagcc	8324
gcacgcgcgtt gatcgccagt ggcagcgcgc cccgctcgat ctggcgcacgc ttctcgatc	8384
ggaaccgggg cgaggccgag gatggcatg acagccacgg cgtccgcgtg gacctgcgt	8444
ggcccggtt ccgcatacgat atcgatggccg tggacaaggg gcccaagcct ttctcgatc	8504
gcgcacggcat gaaccacacg gtcgagacca gcccgtgtt cccgcctgg cctgcgcagg	8564

cggaagcggta	ttgccgcgtc	atcgaggatg	cgatcgccgc	ccgcgcacatg	gccgccttgg	8624
gtccgcgggt	cgaggcgaac	gcccttgcga	tgcacgcccac	gatgatggcc	gcgcgcggc	8684
cgctctgcta	cctgacgggc	ggcagctggc	aggtgcttgg	acgcctgtgg	caggcccccg	8744
cggacgggct	tgcggcctt	gacgatgg	atgcggccc	gaacgtcaag	ctgatcttcg	8804
aggaaagcag	cggccgcac	gtgctgtacc	tgttccccga	cggccgcctg	atcgccgcgt	8864
tcgagggcgt	ttgaacgcgt	aagacgacca	ctgggtaagg	ttctgcccgc	cgtggtctcg	8924
actgcctgca	aagaggtgct	ttagttgtcg	cgtgactgcg	gcggccgact	tcgtggact	8984
tgcccgcac	gctgacgcgc	tggaaacgcg	ccgcggatt	acgaccgcgt	cattgccttg	9044
aaccaatttc	ccgtcggtcg	ac				9066

<210> 43

<211> 340

<212> PRT

<213> Paracoccus sp. R114

<400> 43

Met	Ile	Ser	His	Thr	Pro	Val	Pro	Thr	Gln	Trp	Val	Gly	Pro	Ile	Leu
1					5				10				15		

Phe	Arg	Gly	Pro	Val	Val	Glu	Gly	Pro	Ile	Ser	Ala	Pro	Leu	Ala	Thr
					20			25				30			

Tyr	Glu	Thr	Pro	Leu	Trp	Pro	Ser	Thr	Ala	Arg	Gly	Ala	Gly	Val	Ser
					35			40			45				

Arg	His	Ser	Gly	Gly	Ile	Gln	Val	Ser	Leu	Val	Asp	Glu	Arg	Met	Ser
					50			55			60			65	

Arg	Ser	Ile	Ala	Leu	Arg	Ala	His	Asp	Gly	Ala	Ala	Ala	Thr	Ala	Ala
65					70				75			80			

Trp	Gln	Ser	Ile	Lys	Ala	Arg	Gln	Glu	Val	Ala	Ala	Val	Val	Ala
					85			90			95			

Thr	Thr	Ser	Arg	Phe	Ala	Arg	Leu	Val	Glu	Leu	Asn	Arg	Gln	Ile	Val
					100			105			110				

Gly Asn Leu Leu Tyr Ile Arg Ile Glu Cys Val Thr Gly Asp Ala Ser
115 120 125

Gly His Asn Met Val Thr Lys Ala Ala Glu Ala Val Gln Gly Trp Ile
130 135 140

Leu Ser Glu Tyr Pro Met Leu Ala Tyr Ser Thr Ile Ser Gly Asn Leu
145 150 155 160

Cys Thr Asp Lys Lys Ala Ser Ala Val Asn Gly Ile Leu Gly Arg Gly
165 170 175

Lys Tyr Ala Val Ala Glu Val Glu Ile Pro Arg Lys Ile Leu Thr Arg
180 185 190

Val Leu Arg Thr Ser Ala Glu Lys Met Val Arg Leu Asn Tyr Glu Lys
195 200 205

Asn Tyr Val Gly Gly Thr Leu Ala Gly Ser Leu Arg Ser Ala Asn Ala
210 215 220

His Phe Ala Asn Met Leu Leu Gly Phe Tyr Leu Ala Thr Gly Gln Asp
225 230 235 240

Ala Ala Asn Ile Ile Glu Ala Ser Gln Gly Phe Val His Cys Glu Ala
245 250 255

Arg Gly Glu Asp Leu Tyr Phe Ser Cys Thr Leu Pro Asn Leu Ile Met
260 265 270

Gly Ser Val Gly Ala Gly Lys Gly Ile Pro Ser Ile Glu Glu Asn Leu
275 280 285

Ser Arg Met Gly Cys Arg Gln Pro Gly Glu Pro Gly Asp Asn Ala Arg
290 295 300

Arg Leu Ala Ala Ile Cys Ala Gly Val Val Leu Cys Gly Glu Leu Ser
305 310 315 320

Leu Leu Ala Ala Gln Thr Asn Pro Gly Glu Leu Val Arg Thr His Met
325 330 335

Glu Met Glu Arg
340

<210> 44
<211> 9066
<212> DNA
<213> Paracoccus sp. R114

<220>
<221> CDS
<222> (3641)..(4690)
<223> idi

<400> 44
ggatccggca gctcgacacg ccgcagaacc tgtacgaacg tcccgccagc cgttcgtcg 60
cggaattcgt cgggcgcggg acggtgttgc ccgtgcaggg ccatgacggc gcggggcccg 120
cccgcatcct gggggcccgag gtggcggtga acgcccggcc gcaatcgccgc tttgtcgatc 180
acgtctgcct gcgccccgag aaccttgcca tctccgagac gggcgacctg cgcgccaagg 240
tcgcgcgcgt cacctatctt ggcgggaaat acctgcttga aaccgtgctg gattgcggca 300
cccggttgtt gaccgagacc cgcgcggcgt tcgatacggg cgccgagctt ggcctgacca 360
tcaacgcccc ctgggccttt gccgaggatt gaatggacag cgtgaagatc ctttcggca 420
tgggcgtgaa gggccctgccc tgcatcaggc tggatgtcgg cggatgcgc ctgatcctcg 480
attgcgggac cggccccggac gagggcgcgg agttcgaccc cgcctggctg gcggacgcgg 540
atgcgggtgtt gatcacccat gaccacgtgg accatatcgg cggcgccgt caccggcgtcg 600
cgccggggct gccgatccat gcgacgcggc agacggcggg gttgctgccc gcggggggcgg 660
atctgcgcct gctgcccgaa cgcgggtgtca cgcggatcgc cggggatcgat ctgacgaccg 720
gtcgcaacgg gcatgccgcg ggccgggtct ggatgcattt cgacatggc gaggggctgt 780
tctattccgg cgaactggtcc gaggaatccg actggttcgc cttcgatccg ccccccgcctg 840
cggggacggc gattctcgac tgctcctatg gcgggttcga cgtggcgcaa tcggattgca 900
tcgcggaccc ggacgacctg ctcgaggtgc tgccggggca ggtactgctg cccggcgccgc 960

catccggccg cgccggccgag ctggccctgc ggctgatccg ccgccacgga ccgggcagcg 1020
 tcatggtcga cgacgcctgc ctgccggcca tcgcgcaact gcccggaggcg cgccggactgg 1080
 cctacgccac cgaggcacgc tttcttgcgt gcgcacacgcc gaacgcccga agccggcgcg 1140
 gcatggcggc atctgcaagc atggcgcat gcgggcaggg tggggcggga cgccatgtc 1200
 gtcttcaccg ggcacatgaa cgtccatgcg cgccattct gcgaccgccc cggcgggcat 1260
 ttccgcccgt ggaacgtgca tccgcccgtg cgccgaccagg gacggatgct ggaacggctg 1320
 gccgcgcggc gctttcccccc ggccttctgc cccgaccggc agatctatct ggccgtggac 1380
 atgggcgcgc aggtcttcat gcaccaggag gtgacgcatt gatccccgcg cgccatgtc 1440
 gcctgatccg ccacggcgaa acgaccgcca atgcaggggc gatcatcgcg ggcgcaaccg 1500
 atgtccccct gacgccaagg ggccgcgatc aggcccgcgc cctggcaggg cgccatggc 1560
 catcgggcat cgccgtgttc gccagcccgta tgccgtgtc ccgcgataacc ggcgtgtgg 1620
 cctttccggg ggcgcgaccac cagcccgaaac ccgatctgcg cgaacgcgac tggggcatct 1680
 tcgagggacg ccccgtcgccc gatctcccccc cgccgcgaaat cacgcccgcg gggggcgagg 1740
 gctgggacga cgtgatggcc cgccgtggacc gcgcgatccg gcggatctgc ggcgtgtcg 1800
 gcgcgtgcgt gcgcgtgtcg gtctgccatt cggcggtgtat ccgtgcgcgcg cgccgtgtgt 1860
 ggaccaccgg cgatgcgggc gatcgccgc ccaacgcac gccgatcctg ttccagcccg 1920
 acggcgaccg attaaaggaa ggaacgatat gaccgcccacc acccccgtcg tcgtcttcga 1980
 acgtggacgg cacgcttgcc gaattcgacg ccgaccgcct gggccatctt gtccacggca 2040
 cgaccaagca ctgggacgccc ttccaccacg cgatggccga cgcccccgcctt atccccgagg 2100
 tcgccccctt gatgcgcaag ctgaaggagg ggggcgagac ggtcgatc tgcgtggggc 2160
 gggcccccggg ctggcaggat cagacgatcg catggctgcg caagcacgac ctgccttcg 2220
 acgggatcta tctgcgcccc gaggatcagg acggcgccag cgaccccgag gtcaagcgcc 2280
 gcgccttagc cgagatgcgcg gccgacgggc tggcgccctg gctggtcgtg gacgaccggc 2340
 ggtccgtcgt ggtgcctgg cgggcccggg ggctggtctg cctgcataatgc ggcgggggg 2400
 acttcttaggg cccgcgcacg ggggcgcggc caggctgggc gggaaaaccgc cccgcccacca 2460
 tgtcctgcac gcgtcgaaacc gcccgtccga cgccggtttc cgacacggaaa cgccggcaa 2520
 gttgacataa cttgcacgcg acgtctcgat tctgcggcgc aagaatgcga tgcataccaga 2580
 tgatgcagaa cgaagaagcg gaagcgccccg tgaaagacca gatgatttcc catacccccgg 2640

tgcccacgca atgggtcggc ccgatcctgt tccgcggccc cgtcgtcgag ggccccatca	2700
gcgcgcccgtt ggccacctac gagacgccgc tctggccctc gaccgcgcgg ggggcagggg	2760
tttccccggca ttcgggcggg atccaggtct cgctggtcga cgaacgcattg agccgctcga	2820
tcgcgctgctg ggcgcattgac gggggggcgg cgaccgcgc ctggcagtctg atcaaggccc	2880
gccaggaaga ggtcgccggcc gtggtcgcca ccaccagccg cttcgccgc cttgtcgagc	2940
tgaatcgcca gatcggtggc aacctgcttt acatccgcattt cgaatgcgtg acggggcagc	3000
cctcggtca caacatggtc accaaggccg ccgaggccgt gcagggtctgg atcctgtcgg	3060
aatacccgat gctggcctat tccacagatct cggggAACCT gtgcaccgcac aagaaggcgt	3120
cggcggtcaa cggcatcctg ggccgcggca aatacgcgtt cgccgagggtc gagatcccgc	3180
gcaagatcct gacccgcgtg ctgcgcacca gcgccgagaa gatggtccgc ctgaactacg	3240
agaagaacta tgtcgggggt acgctggcgg ggtcgctgctg cagtgcgaac ggcgcatttcg	3300
ccaacatgtt gctgggcttc tacctggcga cggggcagga cgccggcaac atcatcgagg	3360
ccagccaggg cttcgtccat tgcgaggccc gcggcgagga tctgtatttc tcgtgcacgc	3420
tgccccacct catcatgggc tcggtcggtg cggcaaggg catccccctcg atcgaggaga	3480
acctgtcgctg gatgggctgc cgccagccgg gcgaaccggcg cgacaacgcg cggcgtcttgc	3540
cggcgatctg cgcggcgctc gtgctgtgtg gtgaattgtc gctgcttgcg gcccagacca	3600
accccgagaa gttggtccgc acccacatgg agatggagcg atg acc gac agc aag Met Thr Asp Ser Lys 1 5	3655
 gat cac cat gtc gcg ggg cgc aag ctg gac cat ctg cgt gca ttg gac	3703
Asp His His Val Ala Gly Arg Lys Leu Asp His Leu Arg Ala Leu Asp	
10 15 20	
 gac gat gcg gat atc gac cgg ggc gac agc ggc ttc gac cgc atc gcg	3751
Asp Asp Ala Asp Ile Asp Arg Gly Asp Ser Gly Phe Asp Arg Ile Ala	
25 30 35	
 ctg acc cat cgc gcc ctg ccc gag gtg gat ttc gac gcc atc gac acg	3799
Leu Thr His Arg Ala Leu Pro Glu Val Asp Phe Asp Ala Ile Asp Thr	
40 45 50	
 gcg acc agc ttc ctg ggc cgt gaa ctg tcc ttc ccg ctg atc tcg	3847
Ala Thr Ser Phe Leu Gly Arg Glu Leu Ser Phe Pro Leu Leu Ile Ser	
55 60 65	
 tcc atg acc ggc ggc acc ggc gag gag atc gag cgc atc aac cgc aac	3895
Ser Met Thr Gly Gly Thr Gly Glu Glu Ile Glu Arg Ile Asn Arg Asn	
70 75 80 85	
 ctg gcc gct ggt gcc gag gag gcc cgc gtc gcc atg gcg gtg ggc tcg	3943

Leu Ala Ala Gly Ala Glu Glu Ala Arg Val Ala Met Ala Val Gly Ser			
90	95	100	
cag cgc gtg atg ttc acc gac ccc tcg gcg cggttgcgacctggccatcgt		3991	
Gln Arg Val Met Phe Thr Asp Pro Ser Ala Arg Ala Ser Phe Asp Leu			
105	110	115	
cgc gcc cat gcg ccc acc gtg ccg ctg ctg gcc aat atc ggc gcg gtg		4039	
Arg Ala His Ala Pro Thr Val Pro Leu Leu Ala Asn Ile Gly Ala Val			
120	125	130	
cag ctg aac atg ggg ctg ggg ctg aag gaa tgc ctg gcc gcg atc gag		4087	
Gln Leu Asn Met Gly Leu Gly Leu Lys Glu Cys Leu Ala Ala Ile Glu			
135	140	145	
gtg ctg cag gcg gac ggc ctg tat ctg cac ctg aac ccc ctg caa gag		4135	
Val Leu Gln Ala Asp Gly Leu Tyr Leu His Leu Asn Pro Leu Gln Glu			
150	155	160	165
gcc gtc cag ccc gag ggg gat cgc gac ttt gcc gat ctg ggc agc aag		4183	
Ala Val Gln Pro Glu Gly Asp Arg Asp Phe Ala Asp Leu Gly Ser Lys			
170	175	180	
atc gcg gcc atc gcc cgc gac gtt ccc gtg ccc gtc ctg ctg aag gag		4231	
Ile Ala Ala Ile Ala Arg Asp Val Pro Val Pro Val Leu Leu Lys Glu			
185	190	195	
gtg ggc tgc ggc ctg tcg gcg gcc gat atc gcc atc ggg ctg cgc gcc		4279	
Val Gly Cys Gly Leu Ser Ala Ala Asp Ile Ala Ile Gly Leu Arg Ala			
200	205	210	
ggg atc cgg cat ttc gac gtg gcc ggt cgc ggc aca tcc tgg agc		4327	
Gly Ile Arg His Phe Asp Val Ala Gly Arg Gly Thr Ser Trp Ser			
215	220	225	
cggtatcgcccgcccgcccgatgacgacgttgcgcctggccatcgt		4375	
Arg Ile Glu Tyr Arg Arg Gln Arg Ala Asp Asp Asp Leu Gly Leu			
230	235	240	245
gtc ttc cag gac tgg ggc ctg cag acc gtg gac gcc ctg cgc gag gcg		4423	
Val Phe Gln Asp Trp Gly Leu Gln Thr Val Asp Ala Leu Arg Glu Ala			
250	255	260	
cgg ccc gcg ctt gcg gcc cat gat gga acc agc gtg ctg atc gcc agc		4471	
Arg Pro Ala Leu Ala Ala His Asp Gly Thr Ser Val Leu Ile Ala Ser			
265	270	275	
ggc ggc atc cgc aac ggt gtc gac atg gcg aaa tgc gtc atc ctg ggg		4519	
Gly Gly Ile Arg Asn Gly Val Asp Met Ala Lys Cys Val Ile Leu Gly			
280	285	290	
gcc gac atg tgc ggg gtc gcc gcgc ccc ctg ctg aaa gcg gcc caa aac		4567	
Ala Asp Met Cys Gly Val Ala Ala Pro Leu Leu Lys Ala Ala Gln Asn			
295	300	305	
tcg cgc gag gcg gtt gta tcc gcc atc cgg aaa ctg cat ctg gag ttc		4615	
Ser Arg Glu Ala Val Val Ser Ala Ile Arg Lys Leu His Leu Glu Phe			

310	315	320	325	
cgg aca gcc atg ttc ctc ctg ggt tgc ggc acg ctt gcc gac ctg aag				4663
Arg Thr Ala Met Phe Leu Leu Gly Cys Gly Thr Leu Ala Asp Leu Lys				
330	335	340		
gac aat tcc tcg ctt atc cgt caa tga aagtgcctaa gatgaccgtg				4710
Asp Asn Ser Ser Leu Ile Arg Gln				
345				
acaggaatcg aagcgatcgat cttctacacc ccccagaact acgtgggact ggatatcctt				4770
gccgcgcatac acgggatcga ccccgagaag ttctcgaaagg ggatcgccca ggagaaaaatc				4830
gcactgcccgc gccatgacga ggatatcgta accatggccg ccgaggccgc gctgccgatc				4890
atcgaacgcg cgggcacgca gggcatcgac acggttctgt tcgcccaccga gagcgggatc				4950
gaccagtcga aggccgcccgc catctatctg cgccgcctgc tggacctgtc gcccaactgc				5010
cgttgcgtcg agctgaagca ggcctgctat tccgcgacgg cggcgctgca gatggcctgc				5070
ggcgcattgtcg cccgcaagcc cgaccgcgaa gtgctggtga tcgcgtccga tgtcgccgccc				5130
tatgaccgcg aaagctcggg cgagggcgacg cagggtgcgg gcgcgcgtcg catccttgc				5190
agcgccgatc ccaagggtggc cgagatcgcc accgtctcg ggctgttcac cgaggatatc				5250
atggatttct ggccggccgaa ccaccgcgc acgccccctgt tcgacggcaa ggcacatcgacg				5310
ctgcgtatc tgaacgcgct ggtcgaggcg tggAACGACT atcgcgcgaa tggcggccac				5370
gagttcgccg atttcgcgca tttctgctat cacgtgcgt tctcgccgat gggcgagaag				5430
gcgaacagcc acctggccaa ggcgaacaag acgcccggg acatggggca ggtgcagacg				5490
ggcctgatct acaaccggca ggtcgaaaac tgctataccg ggtcgatcta cctggcattc				5550
gcctcgatgc tggagaacgc tcaggaggac ctgaccggcg cgctggtcgg tctgttcagc				5610
tatggctcgg gtgcgacggg cgaatttttc gatgcgcgga tcgcgcggg ttaccgcgac				5670
cacctgttcg cggAACGCCA tcgcgaattt ctgcaggatc gcacgccccgt cacatatgac				5730
gaatacgttg ccctgtggga cgagatcgac ctgacgcagg gcgcgcggca caaggcgcgc				5790
ggtcgtttca ggctggcagg tatcgaggac gagaagcgca tctatgtcga ccggcaggcc				5850
tgaaggcaggc gcccattgccc cgggcaagct gatcctgtcc gggaaacatt ccgtgctcta				5910
tggcgtcgccc ggcgttgcca tggccatcgcc cgctataacc gaggtgttgt tcacgcgcgt				5970
tggcattggc gaggggatac gcacgacatt cgccaatctc tcggcggggg cgacccatttc				6030
gctgaagctg ctgtcggggt tcaagtcgcg gctggaccgc cggttcgagc agttcctgaa				6090
cgccgaccta aaggtgcaca aggtcctgac ccattccgac gatctggcgg tctatgcgcgt				6150

ggcgtcgctt ctgcacgaca agccgccggg gaccgcccgg atgcccggca tcggcgcgat 6210
 gcaccacactg ccgcgaccgg gtgagctggg cagccggacg gagctgcccc tcggcgcggg 6270
 catggggtcg tctgcggcca tcgtcgccgc caccacggtc ctgttcgaga cgctgctgga 6330
 ccggcccaag acgcccgaac agcgcttcga ccgcgtccgc ttctgcgagc ggttgaagca 6390
 cggaaggcc ggtcccatcg acgcggccag cgtcgtgcgc ggcgggcttg tccgcgtggg 6450
 cggaaacggg ccgggttcga tcagcagctt cgatttgcgg gaggatcacg accttgtcgc 6510
 gggacgcggc tggtaactggg tactgcacgg gcgcggccgc agcgggaccg gcgaatgcgt 6570
 cagcgcggtc gcggcggcgc atggtcgcga tgcggcgctg tgggacgcct tcgcagtctg 6630
 cacccgcgcg ttggaggccg cgctgctgtc tggggcagc cccgacgcgg ccataccga 6690
 gaaccagcgc ctgctggaac gcatcggcgt cgtgcggca gcgacgcagg ccctcgtggc 6750
 ccagatcgag gaggcgggtg gcgcggccaa gatctgcggc gcaggttccg tgcggggcga 6810
 tcacggcggg gcggtcctcg tgcggattga cgacgcgcag gcgatggctt cggatggc 6870
 ggcgcattcc gacctcgact gggcgcctt ggcgcatttcg cgcacggggg cggcacccgg 6930
 cccgcgcgc cgtgcgcac acgcgcggg gcagggttgc tggatcaggt catccgcgc 6990
 agcgcgcgg gttcggtcat gatcacggc gaacatgcgg tggatcgatgg acaccgcgc 7050
 atcgtcgccg ggatcgagca gcgcgcctt gtgcgcgcac tccgcgtgc cgaccgcatt 7110
 tttcgcatca cctcgcatcgat cggggcgcgc cagcagggtt cgctggacga tctgcctgcg 7170
 ggcgggaccc atcgcttcgt gctggccgc atcgcgac acgcgcggg cctgccttc 7230
 gggatcgaca tggacatcac ctggggatc gatccgagggc tcgggcttgg atcctcgccg 7290
 gcgggtgacgg tcgcctgcct cggcgccgtc tcgcggctgg cggggcgggg gaccgagggg 7350
 ctgcattgcg acgcgcgtcg catcgccgc gccatccagg gcaggggcag cggggccat 7410
 ctggcggcca gcctgcattgg cggcttcgtc gcctatcgcc gcggcgatgg cggtgcgcg 7470
 cagatcgagg cgcttcgggt gcccggggg ccgttcggcc tgcgtatgc gggctacaag 7530
 accccgacag ccgaggtgct gcgccttgcg gccgatcgga tggcgggcaa cgaggccgt 7590
 ttgcacgcgc tctactcccg gatggcgca agcgacatgc cgcgcgtccg cgcggcgca 7650
 gggctggact gggctgcatt ccacgacgcg ctgaacgaat accagcgccct gatggagcag 7710
 ctggcgtgt ccgacgacac gctggacgcg atcatccgcg aggcgccgaa cgcggcgcc 7770
 gcagtcgcca agatctccgg ctggggctg gggattgcg tgctggact gggcgaccag 7830

cccaaagggtt tcgtgcccgc aagcattgcc gagaaggac ttgtttcga tgactgatgc 7890
 cgtccgcgac atgatcgccc gtgcctatggc gggcgcgacc gacatccgag cagccgaggc 7950
 ttatgcgccc agcaacatcg cgctgtcgaa atactggggc aagcgcgacg ccgcgcggaa 8010
 ctttcgctg aacagctccg tctcgatctc gttggcgaac tggggctctc atacgcgggt 8070
 cgaggggtcc ggcacgggccc acgacgaggt gcatcacaac ggcacgctgc tggatccggg 8130
 cgacgccttc ggcgcggcg cggtggcatt cgctgacctg ttccgggggg ggaggcacct 8190
 gccgctgcgg atcacgacgc agaactcgat cccgacggcg gcggggcttg cctcgtcggc 8250
 ctcggggttc gccccgtga cccgtgcgtc ggcggggcg ttccggctgg atctggacga 8310
 cacggatctg agccgcacatcg cccggatcgg cagtggcagc gcccggcgct cgatctggca 8370
 cggcttcgtc cgctggaaacc gggggcgaggc cgaggatggg catgacagacc acggcggtccc 8430
 gctggacctg cgctggcccg gcttccgcat cgcatcgatg gccgtggaca agggggcccaa 8490
 gcctttcagt tcgcgcgacg gcatgaacca cacggtcgag accagccgc tggtcccgcc 8550
 ctggccctgca caggcggaag cggattgccg cgtcatcgag gatgcgatcg ccgccccgca 8610
 catggccgccc ctgggtccgc ggtcgagggc gaacgcccctt gcatgcacg ccacgatgat 8670
 ggccgcgcgc ccgcccgtct gctacctgac gggcggcagc tggcagggtgc tggAACGcct 8730
 gtggcaggcc cgcgcgacg ggcttgcggc ctttgcgacg atggatgccg gcccgaacgt 8790
 caagctgatc ttcgagggaaa gcagcgccgc cgacgtgctg tacctgttcc ccgacgcccag 8850
 cctgatcgcc cggttcgagg ggcgttgaac gctgtggaaa cgccggcgcc gattacgacc 8910
 cgcgcggtt ctcgactgcc tgcaaagagg tgcttgagtt gctgcgtgac tgcggcgccc 8970
 gacttcgtgg gacttgcggcc ccacgctgac gctgtggaaa cgccggcgcc gattacgacc 9030
 gcgtcattgc cctgaaccaa ttcccgtcg gtcgac 9066

<210> 45

<211> 349

<212> PRT

<213> Paracoccus sp. R114

<400> 45

Met	Thr	Asp	Ser	Lys	Asp	His	His	Val	Ala	Gly	Arg	Lys	Leu	Asp	His
1				5				10				15			

Leu Arg Ala Leu Asp Asp Asp Ala Asp Ile Asp Arg Gly Asp Ser Gly
20 25 30

Phe Asp Arg Ile Ala Leu Thr His Arg Ala Leu Pro Glu Val Asp Phe
35 40 45

Asp Ala Ile Asp Thr Ala Thr Ser Phe Leu Gly Arg Glu Leu Ser Phe
50 55 60

Pro Leu Leu Ile Ser Ser Met Thr Gly Gly Thr Gly Glu Glu Ile Glu
65 70 75 80

Arg Ile Asn Arg Asn Leu Ala Ala Gly Ala Glu Glu Ala Arg Val Ala
85 90 95

Met Ala Val Gly Ser Gln Arg Val Met Phe Thr Asp Pro Ser Ala Arg
100 105 110

Ala Ser Phe Asp Leu Arg Ala His Ala Pro Thr Val Pro Leu Leu Ala
115 120 125

Asn Ile Gly Ala Val Gln Leu Asn Met Gly Leu Gly Leu Lys Glu Cys
130 135 140

Leu Ala Ala Ile Glu Val Leu Gln Ala Asp Gly Leu Tyr Leu His Leu
145 150 155 160

Asn Pro Leu Gln Glu Ala Val Gln Pro Glu Gly Asp Arg Asp Phe Ala
165 170 175

Asp Leu Gly Ser Lys Ile Ala Ala Ile Ala Arg Asp Val Pro Val Pro
180 185 190

Val Leu Leu Lys Glu Val Gly Cys Gly Leu Ser Ala Ala Asp Ile Ala
195 200 205

Ile Gly Leu Arg Ala Gly Ile Arg His Phe Asp Val Ala Gly Arg Gly
210 215 220

Gly Thr Ser Trp Ser Arg Ile Glu Tyr Arg Arg Arg Gln Arg Ala Asp
225 230 235 240

Asp Asp Leu Gly Leu Val Phe Gln Asp Trp Gly Leu Gln Thr Val Asp
245 250 255

Ala Leu Arg Glu Ala Arg Pro Ala Leu Ala Ala His Asp Gly Thr Ser
260 265 270

Val Leu Ile Ala Ser Gly Gly Ile Arg Asn Gly Val Asp Met Ala Lys
275 280 285

Cys Val Ile Leu Gly Ala Asp Met Cys Gly Val Ala Ala Pro Leu Leu
 290 295 300

Lys Ala Ala Gln Asn Ser Arg Glu Ala Val Val Ser Ala Ile Arg Lys
305 310 315 320

Leu His Leu Glu Phe Arg Thr Ala Met Phe Leu Leu Gly Cys Gly Thr
 325 330 335

Leu Ala Asp Leu Lys Asp Asn Ser Ser Leu Ile Arg Gln
340 345

<210> 46

<211> 9066

<212> DNA

<213> Paracoccus sp. R114

<220>

<221> CDS

<222> (4687) . . (5853)

<223> hcs

```
<400> 46 gatatccggca gctcgacacg ccgcagaacc tgtacgaacg tcccgccagc cgcttcgtcg 60  
cggaattcgt cgggcgcggg acggtggtgc ccgtgcaggc ccatgacggc gcgggccg 120  
ccccatcct gggggccgag gtggcggtga acgccgcccc gcaatcg 180  
acgtctgcct gcgcggcgag aaccttgcca tctccgagac gggcgacctg cgcgccaagg 240
```

tcgcgcgcgt cacctatctt ggcggaaat acctgctgga aaccgtgctg gattgcggca	300
cccggtggt gaccgagacc cgcccccgt tcgatacggg cgcgcagctt ggcctgacca	360
tcaaccccc ctgggccttt gccgaggatt gaatggacag cgtgaagatc ctttcgggca	420
tgggcgtgaa gggccctgcc tgcattcaggc tggatgtcgg cgggatgcgc ctgatcctcg	480
attgcgggac cggcccgac gagggcgcgg agttcgaccc cgcctggctg gcggacgcgg	540
atgcggtgct gatcacccat gaccacgtgg accatatcg cggcgccgt cacgcggtcg	600
cggcggggct gccgatccat gcgacgcggc agacggcggg gttgctgcc gcggggcgg	660
atctgcgcct gctgcccga a cgcggtgtca cgcggatcgc cggggtcgat ctgacgaccg	720
gtcgcaacgg gcatgccgcg ggcggcgtct ggatgcattt cgacatggc gaggggctgt	780
tctattccgg cgactggtcc gaggaatccg actggttcgc cttcgatccg ccccccgcctg	840
cggggacggc gattctcgac tgctcctatg gcggttgcga cgtggcgcaa tcggattgca	900
tcgcggacct ggacgacctg ctcgaggtgc tgccggggca ggtactgctg ccggtgccgc	960
catccggccg cgcggccgag ctggccctgc ggctgatccg cccgcacggc cggggcagcg	1020
tgatggtcga cgacgcctgc ctgcccggca tcgcgcaact gcccggaggcg cgcggactgg	1080
cctacgcccac cgagggcacgc ttcttgcgtc ggcacacgcc gaaacgcccga agccggcgcg	1140
gcatggcggc atctgcaagc atggcgcat ggcggcaggc tggggcggga cgcgcattgtc	1200
gtcttcaccg ggcacatgaa cgtccatgcg cgcgcatttc ggcacccgc cggcgggcat	1260
ttccgcgcgt ggaacgtgca tccgcgcgtg cgcgaccagc gacggatgct ggaacggctg	1320
gccgcgcggc gctttgcggc ggccttctgc cccgaccccg agatctatct ggcgcgtggac	1380
atgggcgcgc aggtcttcat gcaccaggag gtgacgcccattt gatccccgcc cgcagcttct	1440
gcctgatccg ccacggcgaa acgaccgcca atgcaggggc gatcatcgcg ggcgcacccg	1500
atgtccccct gacgccaagg ggccgcgcgc agggccgcgc cctggcaggc cgcgaatggc	1560
catcgggcat cgcgcgttgc gcaaggccga tgtcgcgtgc cgcgcatacc ggcgcgtgg	1620
cctttccggc ggcgcaccac cagcccgaaac ccgatctgcg cgaacgcgcac tggggcatct	1680
tcgagggacg ccccgccgc gatctggccc cgcgcgaaat cacgcgcag gggggcggagg	1740
gctgggacga cgtgatggcc cgcgtggacc ggcgcgcgc ggcgcgttc ggcgcacccgg	1800
gcgcgcgt ggcggcgtcgt gtctgcattt cgggcgtgat ccgtgcgcgc cgcgtgtgt	1860
ggaccacccgg cgcacgcggc gatcgccgc ccaacgcac ggcgcacccgg ttcagcccg	1920

acggcgaccg attaaaggaa ggaacgatat gaccgccacc accccctgcg tcgtttcga	1980
acgtggacgg cacgcttgcg gaattcgacg ccgaccgcct gggccatctt gtccacggca	2040
cгaccaagca ctgggacgcc ttccaccacg cgatggccga cgccccgccc atccccgagg	2100
tсgcccgcct gatgcgcaag ctgaaggagg ggggcgagac ggtcgtcate tgctcggggc	2160
ggccccgccc ctggcaggat cagacgatcg catggctgcg caagcacgac ctgcccattcg	2220
acgggatcta tctgcgcccc gaggatcagg acggcgccag cgaccccgag gtcaagcgcc	2280
gсgcccctagc cgagatgcgc gccgacgggc tggcgccctg gctggtcgtg gacgaccggc	2340
ggtccgtcgt ggatgcctgg cgggcccagg ggctggcttg cctgcaatgc gcgccggggg	2400
acttcttaggg ccgcgcgacg ggggcgcgga caggctggc gggaaaccgc cccgccacca	2460
tgtcctgcac gcgtcgaacc gcccgtccga cgccggtttc cgacggaaa cgcgccggcaa	2520
gttgacataa cttgcacgcg acgtctcgat tctgcccgcg aagaatgcga tgcacccaga	2580
tgatgcagaa cgaagaagcg gaagcgccc taaaagacca gatgattcc catacccccgg	2640
tgcccacgca atgggtcggc ccgatctgt tccgcggccc cgctggtcgag ggcccgatca	2700
gсgcccgcct ggccacctac gagacgcgc tctggccctc gaccgcgcgg gggcagggg	2760
tttcccgcca ttccggcggg atccaggtct cgctggtcga cgaacgcattg agccgctcga	2820
tcgcgtcgcg ggccatgac gggccggcgg cgaccgcgc ctggcagtgc atcaaggccc	2880
gccaggaaga ggtcgccggc gtggtcgcca ccaccagccg cttcgccgc cttgtcgagc	2940
tgaatcgcca gatcggtggc aacctgttt acatccgcat cgaatgcgtg acgggcgacg	3000
cctcgggtca caacatggtc accaaggccg ccgaggccgt gcagggctgg atcctgtcgg	3060
aatacccgat gctggcctat tccacgatct cggggAACCT gtgcacccgac aagaaggcgt	3120
cggcggtcaa cggcatcctg ggccgcggca aatacgccgt cgccgaggtc gagatcccgc	3180
gcaagatcct gacccgcgtg ctgcgcacca gcccgcgagaa gatggtcgc ctgaactacg	3240
agaagaacta tgtcgggggt acgctggcgg ggtcgctgcg cagtgcgaac ggcatttcg	3300
ccaacatgct gctggcttc tacctggcga cggggcagga cgccgcggc acatcgagg	3360
ccagccaggg cttcgccat tgcgaggccc gcggcggagga tctgtatttc tcgtgcacgc	3420
tgccccaccc catcatgggc tcggtcggtg cggcaagg catccccctcg atcgaggaga	3480
acctgtcgcg gatggctgc cgccagccgg gcaacccgg cgacaacgcg cgccgtcttgc	3540
cggcgatctg cgccggcggtc gtgctgtgtg gtgaattgtc gctgcttgcg gcccagacca	3600
accccgggaga gttggtccgc acccacatgg agatggagcg atgaccgaca gcaaggatca	3660

ccatgtcgcg gggcgcaagc tggaccatct gcgtgcattg gacgacgatg cggatatcga	3720
ccggggcgac agcggcttcg accgcacatgc gctgacccat cgcccccgtc ccgagggtgga	3780
tttcgacgcc atcgacacgg cgaccagctt cctggccgt gaactgtcct tccccgtgct	3840
gatctcggtcc atgaccggcg gcaccggcgaa ggagatcgag cgcatcaacc gcaacctggc	3900
cgctggtgcc gaggaggccc gcgtgcccattt ggcgggtggc tcgcagcgctgatgttac	3960
cgaccctcg grgcgggcca gcttcgacccat gcgcaccat ggcggccatcgatgttac	4020
ggccaaatatc ggcgcggtgc agctgaacat ggggctgggg ctgaaggaat gcctggccgc	4080
gatcgaggtg ctgcaggcgg acggcctgta tctgcacctg aacccctgc aagaggccgt	4140
ccagcccgag ggggatcgcg actttgccga tctggcagc aagatcgcgccatcgatgttac	4200
cgacgttccc gtgcccgtcc tgctgaagga ggtggctgc ggcctgtcgg cggccgatata	4260
cgccatcggtccctgcgg gcatccggca ttgcacgtg gccggcgccg gccggcacatc	4320
ctggagccgg atcgagttatc gccggcccca gcggccgat gacgacctgg gcctggcttt	4380
ccaggactgg ggcctgcaga ccgtggacgc cctgcgcgag ggcggccccc cgcttgcggc	4440
ccatgatgga accagcgtgc tgatgcgcag cggccgcattt cgcaacggtg tcgacatggc	4500
gaaatgcgtc atcctggggg ccgacatgtg cgggtcgcc ggcggccatcgatgttac	4560
ccaaaactcg cgcgaggcgg ttgtatccgc catccggaaa ctgcacatcgatgttac	4620
agccatgttc ctccctgggtt gcggcacgtt tgccgacccatcgatgttac	4680
ccgtca atg aaa gtg cct aag atg acc gtg aca gga atc gaa gcg atc Met Lys Val Pro Lys Met Thr Val Thr Gly Ile Glu Ala Ile	4728
1 5 10	
agc ttc tac acc ccc cag aac tac gtg gga ctg gat atc ctt gcc gcg Ser Phe Tyr Thr Pro Gln Asn Tyr Val Gly Leu Asp Ile Leu Ala Ala	4776
15 20 25 30	
cat cac ggg atc gac ccc gag aag ttc tcg aag ggg atc ggg cag gag His His Gly Ile Asp Pro Glu Lys Phe Ser Lys Gly Ile Gly Gln Glu	4824
35 40 45	
aaa atc gca ctg ccc ggc cat gac gag gat atc gtg acc atg gcc gcc Lys Ile Ala Leu Pro Gly His Asp Glu Asp Ile Val Thr Met Ala Ala	4872
50 55 60	
gag gcc gcg ctg ccg atc atc gaa cgc gcg ggc acg cag ggc atc gac Glu Ala Ala Leu Pro Ile Ile Glu Arg Ala Gly Thr Gln Gly Ile Asp	4920
65 70 75	
acg gtt ctg ttc gcc acc gag agc ggg atc gac cag tcg aag gcc gcc Thr Val Leu Phe Ala Thr Glu Ser Gly Ile Asp Gln Ser Lys Ala Ala	4968

80	85	90	
gcc atc tat ctg cgc cgc ctg gac ctg tcg ccc aac tgc cgt tgc Ala Ile Tyr Leu Arg Arg Leu Leu Asp Leu Ser Pro Asn Cys Arg Cys	95 100	105	5016 110
gtc gag ctg aag cag gcc tgc tat tcc gcg acg gcg gcg ctg cag atg Val Glu Leu Lys Gln Ala Cys Tyr Ser Ala Thr Ala Ala Leu Gln Met	115	120	5064 125
gcc tgc gcg cat gtc gcc cgc aag ccc gac cgc aag gtg ctg gtg atc Ala Cys Ala His Val Ala Arg Lys Pro Asp Arg Lys Val Leu Val Ile	130	135	5112 140
gcu tcc gat gtc gcg cgc tat gac cgc gaa agc tcg ggc gag gcg acg Ala Ser Asp Val Ala Arg Tyr Asp Arg Glu Ser Ser Gly Glu Ala Thr	145	150	5160 155
cag ggt gcg ggc gcc gtc gcc atc ctt gtc agc gcc gat ccc aag gtg Gln Gly Ala Gly Ala Val Ala Ile Leu Val Ser Ala Asp Pro Lys Val	160	165	5208 170
gcc gag atc ggc acc gtc tcg ggg ctg ttc acc gag gat atc atg gat Ala Glu Ile Gly Thr Val Ser Gly Leu Phe Thr Glu Asp Ile Met Asp	175	180	5256 185
ttc tgg cgg ccg aac cac cgc cgc acg ccc ctg ttc gac ggc aag gca Phe Trp Arg Pro Asn His Arg Arg Thr Pro Leu Phe Asp Gly Lys Ala	195	200	5304 205
tcg acg ctg cgc tat ctg aac gcg ctg gag gcg tgg aac gac tat Ser Thr Leu Arg Tyr Leu Asn Ala Leu Val Glu Ala Trp Asn Asp Tyr	210	215	5352 220
cgc gcg aat ggc ggc cac gag ttc gcc gat ttc gcg cat ttc tgc tat Arg Ala Asn Gly Gly His Glu Phe Ala Asp Phe Ala His Phe Cys Tyr	225	230	5400 235
cac gtg ccg ttc tcg cgg atg ggc gag aag gcg aac agc cac ctg gcc His Val Pro Phe Ser Arg Met Gly Glu Lys Ala Asn Ser His Leu Ala	240	245	5448 250
aag gcg aac aag acg ccg gtg gac atg ggg cag gtg cag acg ggc ctg Lys Ala Asn Lys Thr Pro Val Asp Met Gly Gln Val Gln Thr Gly Leu	255	260	5496 265
atc tac aac ccg cag gtc ggg aac tgc tat acc ggg tcg atc tac ctg Ile Tyr Asn Arg Gln Val Gly Asn Cys Tyr Thr Gly Ser Ile Tyr Leu	275	280	5544 285
gca ttc gcc tcg ctg gag aac gct cag gag gac ctg acc ggc gcg Ala Phe Ala Ser Leu Leu Glu Asn Ala Gln Glu Asp Leu Thr Gly Ala	290	295	5592 300
ctg gtc ggt ctg ttc agc tat ggc tcg ggt gcg acg ggc gaa ttc ttc Leu Val Gly Leu Phe Ser Tyr Gly Ser Gly Ala Thr Gly Glu Phe Phe	305	310	5640 315

gat gcg cgg atc gcg ccc ggt tac cgc gac cac ctg ttc gcg gaa cgc Asp Ala Arg Ile Ala Pro Gly Tyr Arg Asp His Leu Phe Ala Glu Arg 320 325 330	5688
cat cgc gaa ttg ctg cag gat cgc acg ccc gtc aca tat gac gaa tac His Arg Glu Leu Leu Gln Asp Arg Thr Pro Val Thr Tyr Asp Glu Tyr 335 340 345 350	5736
gtt gcc ctg tgg gac gag atc gac ctg acg cag ggc gcg ccc gac aag Val Ala Leu Trp Asp Glu Ile Asp Leu Thr Gln Gly Ala Pro Asp Lys 355 360 365	5784
gcg cgc ggt ttc agg ctg gca ggt atc gag gac gag aag cgc atc Ala Arg Gly Arg Phe Arg Leu Ala Gly Ile Glu Asp Glu Lys Arg Ile 370 375 380	5832
tat gtc gac cgg cag gcc tga agcaggcgcc catgccccgg gcaagctgat Tyr Val Asp Arg Gln Ala 385	5883
cctgtccggg gaacattccg tgctctatgg tgccgcgcg cttgccatgg ccatgcggc ctataccgag gtgtggttca cgccgcgttg cattggcgag gggatacgca cgacattcgc caatctctcg ggccccggca cctattcgct gaagctgctg tcggggttca agtcgcggct ggaccgcggg ttcgagcagt tcctgaacgg cgacctaaag gtgcacaagg tcctgaccca tcccgcacatcg ctggcggtct atgcgcgtggc gtcgcgtctg cacgacaagg cgccggggac cgccgcgatg ccgggcacatcg ggcgcgtgc ccacctgccc cgaccgggtg agctggcag ccggacggag ctgcccacatcg ggcgcggcat ggggtcgctc gggccacatcg tcgcggccac cacggtcctg ttcgagacgc tgctggaccg gccaagacg cccgaacacg gttcgaccg cgtccgcgttc tgcgagcggt tgaagcacgg caaggccgtt cccatcgacg cggccagcgt cgtgcgcggc gggcttgtcc gcgtggcgcc gAACGGCCG gttcgatca gcagcttcga tttgcggag gatcacgacc ttgtcgccgg acgcggctgg tactgggtac tgacacggcgc ccccgtcagc gggaccggcg aatgcgtcag cgccgcgtcg cgccgcgtat gtcgcgtatgc ggcgctgtgg gacgccttcg cagtcgtcac ccgcgcgttg gaggccgcgc tgctgtctgg gggcagcccc gacgcccgcac tcaccgagaa ccagcgcctg ctggAACGCA tcggcgctgt gccggcagcg acgcaggccc tcgtggccca gatcgaggag gcccgtggcg cggccaaagat ctgcggcgca ggttccgtgc gggcgatca cggccggccg gtcctcgatc ggattgacga cgccgcaggcg atggcttcgg tcatggcgccg ccatcccac ctcgactggg cggccctgcg catgtcgccgc acggggggcgcc caccggcccc cgccgcgt ggcacaccgc tgccggggca	5943 6003 6063 6123 6183 6243 6303 6363 6423 6483 6543 6603 6663 6723 6783 6843 6903 6963

gggctgatgg atcaggtcat ccgcgccagc gcgccgggtt cggcatgat cacggcgaa	7023
catgccgtgg tctatggaca ccgcgccatc gtcgcccggga tcgagcagcg cgcccatgtg	7083
acgatcgtcc cgcgtgccga ccgcatttt cgcatcacct cgcaatcg ggccgcgcag	7143
caggggtcgc tggacgatct gcctgcggc gggacctatc gcttcgtgct ggccgcgcac	7203
cgccgacacg cgccggaccc gccttgcggg ttgcacatgg acatcaccc gggatcgat	7263
ccgaggctcg ggcttgatc ctcggcggcg gtgacggtcg cctgcctcgg cgccgtgtcg	7323
cggtggcgg ggccggggac cgaggggctg catgacgacg cgctgcgcac cgtccgcgc	7383
atccaggca gggcagcgg ggccgatctg gcccgcagcc tgcatggcgg ctgcgtgc	7443
tatcgccgc cccatggcgg tgccgcgcag atcgaggcgc ttccggtgcc gccggggccg	7503
ttcggcctgc gctatgcggg ctacaagacc ccgacagccg aggtgctgcg cttgtggcc	7563
gatcgatgg cggcaacga ggccgcattt gacgcgcctactccggat gggcgcaagc	7623
gcagatgccg cgatccgcgc ggcgcaaggg ctggactggg ctgcattcca cgacgcgc	7683
aacgaataacc agcgctgat ggagcagctg ggcgtgtccg acgacacgcg ggacgcgatc	7743
atccgcgagg cgccgcacgc gggcgccgca gtcgccaaga tctccggctc gggctgggg	7803
gattgcgtgc tggcaactggg cgaccagccc aagggttcg tgcccgcaag cattgccgag	7863
aaggacttg tttcgatga ctgatgccgt ccgcgcacatg atcgccgcgt ccatggcggg	7923
cgcgaccgac atccgagcag ccgaggctt gtcgcccagc aacatcgccg tgtcgaaata	7983
ctggggcaag cgccgcacgc cgccgaaacct tccgcgtgaa acgtccgcgt cgatctcg	8043
ggcgaactgg ggctctata cgccggcga ggggtccggc acggggccacg acgaggtgca	8103
tcacaacggc acgctgctgg atccggcga ccgcattcgcc cgccgcgcgt tggcattcg	8163
tgacctgttc cggggggggga ggcacctgccc gctgcggatc acgacgcaga actcgatccc	8223
gacggcggcg gggcttgctt cgtccgcctc ggggttcgac ggcgtgaccc gtgcgtggc	8283
ggggcggttc gggctggatc tggacgacac ggatctgagc cgcatcgccc ggatcggcag	8343
tggcagcgc gccccgtcga tctggcacgg cttcgccgc tggAACGGG gcgaggccga	8403
ggatggccat gacagccacg ggcgtccgcgt ggacctgcgc tggcccggct tccgcacgc	8463
gatcggtggcc tggacacaagg ggcggcaagcc tttcagttcg cgcgcacggca tgaaccacac	8523
ggtcgagacc agcccgctgt tccgcgcgt ggcgcgcag ggcggaaacggg attgcgcgt	8583
catcgaggat ggcgcacat ggcgcgcgtt ggtccgcggg tggaggcgaa	8643
cgcccttgcg atgcacgcac cgcgcgcgcg ccgcgcgtcg acctgacggg	8703

cggcagctgg caggtgctgg aacgcctgtg gcaggccgc gcggacgggc ttgcggcctt	8763
tgcgacgatg gatgccggcc cgaacgtcaa gctgatcttc gaggaaagca gcgccgcca	8823
cgtgctgtac ctgttccccg acgccagcct gatcgccgc ttcgaggggc gttgaacgcg	8883
taagacgacc actgggtaag gttctgccgc gcgtggtctc gactgcctgc aaagaggtgc	8943
ttgagttgct gcgtgactgc ggccggac ttcgtggac ttgcccgcga cgctgacgcg	9003
ctggaaacgc gcccgcggat tacgaccgcg tcattgcct gaaccaattt cccgtcggc	9063
gac	9066

<210> 47

<211> 388

<212> PRT

<213> Paracoccus sp. R114

<400> 47

Met Lys Val Pro Lys Met Thr Val Thr Gly Ile Glu Ala Ile Ser Phe			
1	5	10	15
10	15		

Tyr Thr Pro Gln Asn Tyr Val Gly Leu Asp Ile Leu Ala Ala His His			
20	25	30	
30			

Gly Ile Asp Pro Glu Lys Phe Ser Lys Gly Ile Gly Gln Glu Lys Ile			
35	40	45	
45			

Ala Leu Pro Gly His Asp Glu Asp Ile Val Thr Met Ala Ala Glu Ala			
50	55	60	
60			

Ala Leu Pro Ile Ile Glu Arg Ala Gly Thr Gln Gly Ile Asp Thr Val			
65	70	75	80
75	80		

Leu Phe Ala Thr Glu Ser Gly Ile Asp Gln Ser Lys Ala Ala Ala Ile			
85	90	95	
95			

Tyr Leu Arg Arg Leu Leu Asp Leu Ser Pro Asn Cys Arg Cys Val Glu			
100	105	110	
110			

Leu Lys Gln Ala Cys Tyr Ser Ala Thr Ala Ala Leu Gln Met Ala Cys

115	120	125
Ala His Val Ala Arg Lys Pro Asp Arg Lys Val Leu Val Ile Ala Ser		
130	135	140
Asp Val Ala Arg Tyr Asp Arg Glu Ser Ser Gly Glu Ala Thr Gln Gly		
145	150	155
160		
Ala Gly Ala Val Ala Ile Leu Val Ser Ala Asp Pro Lys Val Ala Glu		
165	170	175
Ile Gly Thr Val Ser Gly Leu Phe Thr Glu Asp Ile Met Asp Phe Trp		
180	185	190
(
Arg Pro Asn His Arg Arg Thr Pro Leu Phe Asp Gly Lys Ala Ser Thr		
195	200	205
Leu Arg Tyr Leu Asn Ala Leu Val Glu Ala Trp Asn Asp Tyr Arg Ala		
210	215	220
Asn Gly Gly His Glu Phe Ala Asp Phe Ala His Phe Cys Tyr His Val		
225	230	235
240		
Pro Phe Ser Arg Met Gly Glu Lys Ala Asn Ser His Leu Ala Lys Ala		
245	250	255
(
Asn Lys Thr Pro Val Asp Met Gly Gln Val Gln Thr Gly Leu Ile Tyr		
260	265	270
(
Asn Arg Gln Val Gly Asn Cys Tyr Thr Gly Ser Ile Tyr Leu Ala Phe		
275	280	285
Ala Ser Leu Leu Glu Asn Ala Gln Glu Asp Leu Thr Gly Ala Leu Val		
290	295	300
Gly Leu Phe Ser Tyr Gly Ser Gly Ala Thr Gly Glu Phe Phe Asp Ala		
305	310	315
320		
Arg Ile Ala Pro Gly Tyr Arg Asp His Leu Phe Ala Glu Arg His Arg		
325	330	335
Glu Leu Leu Gln Asp Arg Thr Pro Val Thr Tyr Asp Glu Tyr Val Ala		
340	345	350

Leu Trp Asp Glu Ile Asp Leu Thr Gln Gly Ala Pro Asp Lys Ala Arg
355 360 365

Gly Arg Phe Arg Leu Ala Gly Ile Glu Asp Glu Lys Arg Ile Tyr Val
370 375 380

Asp Arg Gln Ala
385

<210> 48

<211> 9066

<212> DNA

<213> Paracoccus sp. R114

<220>

<221> CDS

<222> (5834)..(6970)

<223> mvk

<400> 48
ggatccggca gctcgacacg ccgcagaacc tgtacgaacg tccccccagc cgcttcgtcg 60
cggaattcgt cgggcgcggg acggtgtgc ccgtcaggc ccatgacggc gcgggccg 120
cccgcatcct gggggccgag gtggcggtga acgcccggcc gcaatcg 180
acgtctgcct gcgcggcggag aaccttgcca tctccgagac gggcgacctg cgcccaagg 240
tcgcgcgcgt cacctatctt ggcgggaaat acctgctgga aaccgtgctg gattgcggca 300
cccggtggcgt gaccgagacc cgccggcgt tcgatacggg cgccgacgtt ggctgacca 360
tcaacgcccc ctgggccttt gccgaggatt gaatggacag cgtgaagatc ctttcggca 420
tgggcgtgaa gggccctgcc tgcatcaggc tggatgtcg cggtatgcgc ctgatcctcg 480
attgcgggac cggccggac gagggcgcgg agttcgaccc cgccctggctg gcggacgcgg 540
atgcgggtgct gatcacccat gaccacgtgg accatatcg cgccgcgcgt cacgcggctg 600
cgccggggct gccgatccat gcgacgcggc agacggcggg gttgctgccc gcggggcgg 660

atctgcgcct	gctgcccga	cgcggtgtca	cgcggatcgc	cggggtcgat	ctgacgaccg	720
gtcgcaacgg	gcatgcccg	ggcggcgtct	ggatgcattt	cgacatgggc	gaggggctgt	780
tctattccgg	cgactggtcc	gaggaatccg	actggttcgc	cttcgatccg	cccccgctg	840
cggggacggc	gattctcgac	tgctcctatg	gcggttcga	cgtggcgcaa	tcggattgca	900
tcgcccacct	ggacgacctg	ctcgaggtgc	tgccggggca	ggtactgctg	ccgggtgccgc	960
catccggccg	cgcggccgag	ctggccctgc	ggctgatccg	ccgccacgga	ccgggcagcg	1020
tgatggtcga	cgacgcctgc	ctgcccggca	tcgcgcaact	gcccgaggcg	cgcggactgg	1080
cctacccac	cgaggcacgc	tttcttgtct	gcgacacgcc	gaacgcccga	agccggcg	1140
gcatggccgc	atctgcaagc	atggcgcat	gcgggcaggg	tggggcggga	cgcgcatgtc	1200
gtcttcaccg	ggcacatgaa	cgtccatgcg	cgcgcattct	gcgaccgccc	cggcgggcat	1260
ttccggccct	ggaacgtgca	tccggccgtc	cgcgaccagc	gacggatgct	ggaacggctg	1320
gccgcgcggc	gttttgc	ggcctctgc	cccgaccccg	agatctatct	ggcgctggac	1380
atgggcgcgc	aggcttcat	gcaccaggag	gtgacgccc	gatccccgcc	cgcagcttct	1440
gcctgatccg	ccacggcgaa	acgaccgcca	atgcaggggc	gatcatcg	ggcgcaaccg	1500
atgtgcccct	gacgccaagg	ggccgcgatc	aggcccgcgc	cctggcaggg	cgcgaatggc	1560
catcgggcat	cgcgcgttcc	gccagccga	tgtcgctgc	ccgcgatacc	gchgctgtgg	1620
cctttccggg	gcfgcaccac	cagccgaac	ccgatctgcg	cgaacgcgac	tggggcatct	1680
tcgagggacg	ccccgtcgcc	gatctgcccc	cgcgcgaaat	cacgcccgcag	ggggggcgagg	1740
gctggacga	cgtatggcc	cgcgtggacc	gcgcgatccg	gcggatctgc	gcgacccctgg	1800
gcgcgcgcgt	gccgggtctg	gtctgccatt	cggcgtgat	ccgtgcccgcg	cgcgtgtgt	1860
ggaccaccgg	cgcgcgggc	gatgtccgc	ccaacgcccac	gccgatctgc	ttcagcccg	1920
acggcgaccg	attaaaggaa	ggaacgatat	gaccgccacc	accccctgcg	tcgtcttcga	1980
acgtggacgg	cacgcttgcc	gaattcgacg	ccgaccgcct	ggccatctt	gtccacggca	2040
cgaccaagca	ctgggacgcc	ttccaccacg	cgcgcgcga	cgccccgccc	atccccgagg	2100
tcgcccgcct	gatgcgcaag	ctgaaggagg	ggggcgagac	ggtcgtcatc	tgctcggggc	2160
ggccccgcgg	ctggcaggat	cagacgatcg	catggctgcg	caagcacgac	ctgcccctcg	2220
acgggatcta	tctgcgcccc	gaggatcagg	acggcgccag	cgaccccgag	gtcaagcgcc	2280
gcgccttagc	cgcgcgcgc	gccgacgggc	tggcgccctg	gctggtcgt	gacgaccggc	2340
ggtcgcgt	ggatgcctgg	cgggcccagg	ggctggtctg	cctgcaatgc	gcccgggggg	2400

acttctaggg ccgcgcgacg	ggggcgccga caggctggc	ggaaaccgc cccgccacca	2460
tgtcctgcac gcgtcgaacc	gcccgtccga cgccggttc	cgcacggaaa cgcgccgaa	2520
gttgacataa ctgcacgcg acgtctcgat	tctgcccgcg aagaatgcga	tgcatccaga	2580
tgatgcagaa cgaagaagcg	gaagcgcccgt gaaaagacca	gatgatttcc catacccccgg	2640
tgccccacgca atgggtcggc	ccgatcctgt tccgcggccc	cgtcgctcgag ggcccgtatca	2700
gcgcgcgcgt ggccacctac	gagacgcccgc tctggccctc	gaccgcgcgg gggcagggg	2760
tttccggca ttccggcggg atccaggct	cgctggtcga cgaacgcatg	agccgctcgaa	2820
tcgcgctgcg ggcatgac	ggggcggcgg cgaccgcgc	ctggcagtctg atcaaggccc	2880
gccaggaaga ggtcgccggc	gtggtcgcca ccaccagccg	cttcgcccgc cttgtcgagc	2940
tgaatcgcca gatcgtggc	aacctgctt acatccgcat	cgaatgcgtg acgggcgacg	3000
cctcgggtca caacatggtc	accaaggccg ccgaggccgt	gcagggctgg atccgtcgaa	3060
aatacccgat gctggcctat	tccacgatct cggggAACCT	gtgcaccgcac aagaaggcgt	3120
cggcgtcaa cggcatcctg	ggccgcggca aatacggccgt	cgccgagggtc gagatcccgc	3180
gcaagatcct gaccgcgtg	ctgcgcacca gcgcgcgagaa	gatggtccgc ctgaactacg	3240
agaagaacta tgtcgggggt	acgctggcgg ggtcgctcg	cagtgcgaac ggcatttcg	3300
ccaacatgct gctggcttc	tacctggcga cggggcagga	cgccgcacaac atcatcgagg	3360
ccagccaggg cttcgtccat	tgcgaggccc gggcggagga	tctgtatttc tcgtgcacgc	3420
tgcccaacct catcatggc	tcggtcggtg cggcaaggg	catcccctcg atcgaggaga	3480
acctgtcgcg gatgggctgc	cgccagccgg gcaacccgg	cgacaacgcg cgccgtcttgc	3540
cggcgatctg cggggcggtc	gtgctgtgtg gtgaattgtc	gctgcttgcg gcccagacca	3600
accccgagaa gtttgtccgc	acccacatgg agatggagcg	atgaccgaca gcaaggatca	3660
ccatgtcgcg gggcgcaagc	tggaccatct gcgtgcattt	gacgacgatg cggatatcga	3720
ccggggcgac agccgttcg	accgcattgc gctgaccat	cgccgcctgc ccgaggtgga	3780
tttcgacgcc atcgacacgg	cgaccagctt cctggccgt	gaactgtcct tcccgctgt	3840
gatctcgcc atgaccggcg	gcaccggcga ggagatcgag	cgcatcaacc gcaacctggc	3900
cgctggtgcc gaggaggccc	gcgtcgccat ggcggtggc	tcgcagcgcg tgatgttcac	3960
cgaccgcgtc ggcggggcca	gcttcgaccc ggcgcgcct	gcccacccg tgccgctgt	4020
ggccaatatc ggcgcgggtgc	agctgaacat gggctgggg	ctgaaggaat gcctggccgc	4080

gatcgaggtg	ctgcaggcgg	acggcctgta	tctgcacctg	aaccccctgc	aagaggccgt	4140
ccagccccag	ggggatcgcg	actttgccga	tctgggcagc	aagatcgcg	ccatcgcccc	4200
cgacgttccc	gtgcccgtcc	tgctgaagga	ggtgggctgc	ggcctgtcgg	cggccgatat	4260
cgccatcggg	ctgcgcgccc	ggatccggca	tttcgacgtg	gccggtcgcg	gcggcacatc	4320
ctggagccgg	atcgagtatc	gccgcccaca	gcgggcccgt	gacgacctgg	gcctggtctt	4380
ccaggactgg	ggcctgcaga	ccgtggacgc	cctgcgcgag	gcgcggcccg	cgcttgcggc	4440
ccatgatgga	accagcgtgc	tgatcgccag	cggccgcatac	cgcaacggtg	tcgacatggc	4500
gaaatgcgtc	atcctggggg	ccgacatgtg	cggggtcgccc	gcgc(ccctgc	tgaaagcggc	4560
ccaaaactcg	cgcgaggcgg	ttgtatccgc	catccggaaa	ctgcatctgg	agttccggac	4620
agccatgttc	ctcctgggtt	gcggcacgct	tgccgacctg	aaggacaatt	cctcgcttat	4680
ccgtcaatga	aagtgcctaa	gatgaccgtg	acaggaatcg	aagcgatcg	cttctacacc	4740
ccccagaact	acgtgggact	ggatatcctt	gccgcgcatac	acgggatcga	ccccgagaag	4800
ttctcgaagg	ggatcgggca	ggagaaaatc	gcactgccc	gccatgacga	ggatatcg	4860
accatggccg	ccgaggccgc	gctgccgatc	atcgaacgcg	cggcacgca	gggcacatcgac	4920
acggttctgt	tcgcccaccga	gagcgggatc	gaccagtcga	aggccgcgc	catctatctg	4980
cgccgcctgc	tggacctgtc	gcccaactgc	cgttgcgtcg	agctgaagca	ggcctgctat	5040
tccgcgacgg	cggcgctgca	gatggcctgc	gcgcatgtcg	cccgcaagcc	cgaccgcaag	5100
gtgctggtga	tcgcgtccga	tgtcgcgccg	tatgaccgcg	aaagctcggg	cgaggcgacg	5160
cagggtgtcg	gcccgcgtcg	catccttgc	agcgccgatc	ccaagggtggc	cgagatcg	5220
accgtctcg	ggctgttcac	cgaggatatc	atggatttct	ggcggccgaa	ccaccgccc	5280
acgc(ccctgt	tcgacggcaa	ggcatcgacg	ctgcgcatac	tgaacgcgct	ggtcgaggcg	5340
tggaacgact	atcgcgca	tggcggccac	gagttcgccg	atttcgcgca	tttctgctat	5400
cacgtgccgt	tctcgccgat	gggcgagaag	gcgaacagcc	acctggccaa	ggcgaacaag	5460
acgcccgtgg	acatggggca	ggtgcagacg	ggcctgatct	acaaccggca	ggtcggaaac	5520
tgctataccg	ggtcgatcta	cctggcattc	gcctcgctgc	tggagaacgc	tcaggaggac	5580
ctgaccggcg	cgctggtcgg	tctgttcagc	tatggctcgg	gtgcgacggg	cgaattttc	5640
gatgcgcgga	tcgcgcccgg	ttaccgcgac	cacctgttcg	cggaacgcca	tcgcgaattg	5700
ctgcaggatc	gcacgcccgt	cacatatgac	gaatacgttg	ccctgtggga	cgagatcgac	5760
ctgacgcagg	gcccgcgt	caaggcgcgc	ggtcgttca	ggctggcagg	tatcgaggac	5820

gagaagcgca tct atg tcg acc ggc agg cct gaa gca ggc gcc cat gcc Met Ser Thr Gly Arg Pro Glu Ala Gly Ala His Ala 1 5 10	5869
ccg ggc aag ctg atc ctg tcc ggg gaa cat tcc gtg ctc tat ggt gcg Pro Gly Lys Leu Ile Leu Ser Gly Glu His Ser Val Leu Tyr Gly Ala 15 20 25	5917
ccc gcg ctt gcc atg gcc atc gcc cgc tat acc gag gtg tgg ttc acg Pro Ala Leu Ala Met Ala Ile Ala Arg Tyr Thr Glu Val Trp Phe Thr 30 35 40	5965
ccg ctt ggc att ggc gag ggg ata cgc acg aca ttc gcc aat ctc tcg Pro Leu Gly Ile Gly Glu Gly Ile Arg Thr Thr Phe Ala Asn Leu Ser 45 50 55 60	6013
ggc ggg gcg acc tat tcg ctg aag ctg ctg tcg ggg ttc aag tcg cgg Gly Gly Ala Thr Tyr Ser Leu Lys Leu Ser Gly Phe Lys Ser Arg 65 70 75	6061
ctg gac cgc cgg ttc gag cag ttc ctg aac ggc gac cta aag gtg cac Leu Asp Arg Arg Phe Glu Gln Phe Leu Asn Gly Asp Leu Lys Val His 80 85 90	6109
aag gtc ctg acc cat ccc gac gat ctg gcg gtc tat gcg ctg gcg tcg Lys Val Leu Thr His Pro Asp Asp Leu Ala Val Tyr Ala Leu Ala Ser 95 100 105	6157
ctt ctg cac gac aag ccg ccg ggg acc gcc gcg atg ccg ggc atc gcc Leu Leu His Asp Lys Pro Pro Gly Thr Ala Ala Met Pro Gly Ile Gly 110 115 120	6205
gcg atg cac cac ctg ccg cga ccg ggt gag ctg ggc agc cgg acg gag Ala Met His His Leu Pro Arg Pro Gly Glu Leu Gly Ser Arg Thr Glu 125 130 135 140	6253
ctg ccc atc ggc gcg ggc atg ggg tcg tct gcg gcc atc gtc gcg gcc Leu Pro Ile Gly Ala Gly Met Gly Ser Ser Ala Ala Ile Val Ala Ala 145 150 155	6301
acc acg gtc ctg ttc gag acg ctg ctg gac ccg ccc aag acg ccc gaa Thr Thr Val Leu Phe Glu Thr Leu Leu Asp Arg Pro Lys Thr Pro Glu 160 165 170	6349
cag cgc ttc gac cgc gtc cgc ttc tgc gag ccg ttg aag cac ggc aag Gln Arg Phe Asp Arg Val Arg Phe Cys Glu Arg Leu Lys His Gly Lys 175 180 185	6397
gcc ggt ccc atc gac gcg gcc agc gtc gtg cgc ggc ggg ctt gtc cgc Ala Gly Pro Ile Asp Ala Ala Ser Val Val Arg Gly Gly Leu Val Arg 190 195 200	6445
gtg ggc ggg aac ggg ccg ggt tcg atc agc acg ttc gat ttg ccc gag Val Gly Gly Asn Gly Pro Gly Ser Ile Ser Ser Phe Asp Leu Pro Glu 205 210 215 220	6493

gat cac gac ctt gtc gcg gga cgc ggc tgg tac tgg gta ctg cac ggg Asp His Asp Leu Val Ala Gly Arg Gly Trp Tyr Trp Val Leu His Gly 225	230	235	6541	
cgc ccc gtc agc ggg acc ggc gaa tgc gtc agc gcg gtc gcg gcg gcg Arg Pro Val Ser Gly Thr Gly Glu Cys Val Ser Ala Val Ala Ala Ala 240	245	250	6589	
cat ggt cgc gat gcg gcg ctg tgg gac gcc ttc gca gtc tgc acc cgc His Gly Arg Asp Ala Ala Leu Trp Asp Ala Phe Ala Val Cys Thr Arg 255	260	265	6637	
gcg ttg gag gcc gcg ctg ctg tct ggg ggc agc ccc gac gcc gcc atc Ala Leu Glu Ala Ala Leu Ser Gly Gly Ser Pro Asp Ala Ala Ile 270	275	280	6685	
acc gag aac cag cgc ctg ctg gaa cgc atc ggc gtc gtg ccg gca gcg Thr Glu Asn Gln Arg Leu Leu Glu Arg Ile Gly Val Val Pro Ala Ala 285	290	295	300	6733
acg cag gcc ctc gtg gcc cag atc gag gag gcg ggt ggc gcg gcc aag Thr Gln Ala Leu Val Ala Gln Ile Glu Glu Ala Gly Gly Ala Ala Lys 305	310	315	6781	
atc tgc ggc gca ggt tcc gtg cgg ggc gat cac ggc ggg gcg gtc ctc Ile Cys Gly Ala Gly Ser Val Arg Gly Asp His Gly Gly Ala Val Leu 320	325	330	6829	
gtg cgg att gac gac gcg cag gcg atg gct tcg gtc atg gcg cgc cat Val Arg Ile Asp Asp Ala Gln Ala Met Ala Ser Val Met Ala Arg His 335	340	345	6877	
ccc gac ctc gac tgg gcg ccc ctg cgc atg tcg cgc acg ggg gcg gca Pro Asp Leu Asp Trp Ala Pro Leu Arg Met Ser Arg Thr Gly Ala Ala 350	355	360	6925	
ccc ggc ccc gcg ccg cgt gcg caa ccg ctg ccg ggg cag ggc tga Pro Gly Pro Ala Pro Arg Ala Gln Pro Leu Pro Gly Gln Gly 365	370	375	6970	
tggatcaggt catccgcgcc agcgcgcgg gttcggtcat gatcacggc gaacatgccg			7030	
tggcttatgg acaccgcgcc atcgctgccg ggatcgagca gcgcgcctat gtgacgatcg			7090	
tccccgcgtgc cgaccgcatg tttcgcatca cctcgcagat cggggcgccg cagcaggggt			7150	
cgctggacga tctgcctgacg ggcgggacact atcgcttcgt gctggccgc atcgccgcac			7210	
acgcgcgcga cctgccttgc gggttcgaca tggacatcac ctcggggatc gatccgaggc			7270	
tcggggcttgg atcctcggcg gcggtgacgg tcgcctgcct cggcgcgctg tcgcggctgg			7330	
cggggcgaaa gaccgagggg ctgcattgacg acgcgcgtcg catcgccgc gccatccagg			7390	
gcaggggcag cggggccgat ctggcgccca gcctgcattgg cggcttcgtc gcctatcgcg			7450	
cggccgatgg cggtgccgcg cagatcgagg cgcttccgggt gccgcgggg ccgttcggcc			7510	

tgcgctatgc	gggctacaag	accccgacag	ccgagggtct	gcgccttgtg	gccgatcgga	7570
tggcgggcaa	cgaggccgct	ttcgacgcgc	tctactccc	gatgggcgca	agcgcagatg	7630
ccgcgatccg	cgcggcgcaa	gggctggact	gggctgcatt	ccacgacgcg	ctgaacgaat	7690
accagcgct	gatggagcag	ctgggcgtgt	ccgacgacac	gctggacgcg	atcatccgc	7750
aggcgcgcga	cgcgggcgcc	gcagtcgcca	agatctccgg	ctcggggctg	ggggattgct	7810
tgctggcact	gcgcgaccag	ccaaagggtt	tcgtgcccgc	aagcattgcc	gagaagggac	7870
ttgtttcga	tgactgatgc	cgtccgcgac	atgatcgccc	gtgccaatggc	gggcgcgacc	7930
gacatccgag	cagccgagggc	ttatgcgccc	agcaacatcg	cgctgtcgaa	atactggggc	7990
aagcgcgacg	ccgcgcggaa	ccttcgcgtg	aacagctccg	tctcgatctc	gttggcgaac	8050
tggggctctc	atacgcgggt	cgaggggtcc	ggcacgggcc	acgacgaggt	gcatacacaac	8110
ggcacgctgc	tggatccggg	cgacgccttc	gcgcgcgcgc	cgttggcatt	cgctgacctg	8170
ttccgggggg	ggaggcacct	gccgctgcgg	atcacgacgc	agaactcgat	ccgcacggcg	8230
gcggggcttg	cctcgtcggc	ctcggggttc	gcggcgctga	cccggtcgct	ggcgaaaaaa	8290
ttcgggctgg	atctggacga	cacggatctg	agccgcacatcg	cccgatcg	cagtggcagc	8350
gccgcggcgt	cgatctggca	cggcttcgtc	cgctggaaacc	ggggcgagggc	cgaggatggg	8410
catgacagcc	acggcgcccc	gctggacctg	cgctggcccg	gcttcgcacat	cgcgatcg	8470
gccgtggaca	agggggccaa	gccttcagt	tcgcgcacg	gcatgaacca	cacggatcg	8530
accagccgc	tgttcccgcc	ctggcctgcg	caggcggaaag	cggattgccc	cgtcatcg	8590
gatgcgatcg	ccgccccgca	catggccgccc	ctgggtccgc	gggtcgagggc	gaacgcctt	8650
gcgatgcacg	ccacgatgat	ggccgcgcgc	ccgcgcgtct	gctacctgac	gggcggcagc	8710
tggcagggtc	tggaacgcct	gtggcaggcc	cgcgcgacg	ggcttgcggc	ctttgcacg	8770
atggatgccc	gccccaacgt	caagctgatc	ttcgaggaaa	gcagcgccgc	cgacgtgctg	8830
tacctgttcc	ccgacgcccag	cctgatgcgc	ccgttcgagg	ggcggttgaac	gcgttaagacg	8890
accactgggt	aagggttctgc	cgcgcgtgg	ctcgactgcc	tgcaaagagg	tgcttgagtt	8950
gctgcgtgac	tgcggcgccc	gacttcgtgg	gacttgcgg	ccacgctgac	gctgcgtggaaa	9010
cgcgcggcgcg	gattacgacc	gcgtcattgc	cctgaaccaa	tttcccgctg	gtcgac	9066

<210> 49

<211> 378

<212> PRT

<213> Paracoccus sp. R114

<400> 49

Met Ser Thr Gly Arg Pro Glu Ala Gly Ala His Ala Pro Gly Lys Leu
1 5 10 15

Ile Leu Ser Gly Glu His Ser Val Leu Tyr Gly Ala Pro Ala Leu Ala
20 25 30

Met Ala Ile Ala Arg Tyr Thr Glu Val Trp Phe Thr Pro Leu Gly Ile
35 40 45

Gly Glu Gly Ile Arg Thr Thr Phe Ala Asn Leu Ser Gly Gly Ala Thr
50 55 60

Tyr Ser Leu Lys Leu Leu Ser Gly Phe Lys Ser Arg Leu Asp Arg Arg
65 70 75 80

Phe Glu Gln Phe Leu Asn Gly Asp Leu Lys Val His Lys Val Leu Thr
85 90 95

His Pro Asp Asp Leu Ala Val Tyr Ala Leu Ala Ser Leu Leu His Asp
100 105 110

Lys Pro Pro Gly Thr Ala Ala Met Pro Gly Ile Gly Ala Met His His
115 120 125

Leu Pro Arg Pro Gly Glu Leu Gly Ser Arg Thr Glu Leu Pro Ile Gly
130 135 140

Ala Gly Met Gly Ser Ser Ala Ala Ile Val Ala Ala Thr Thr Val Leu
145 150 155 160

Phe Glu Thr Leu Leu Asp Arg Pro Lys Thr Pro Glu Gln Arg Phe Asp
165 170 175

Arg Val Arg Phe Cys Glu Arg Leu Lys His Gly Lys Ala Gly Pro Ile
180 185 190

Asp Ala Ala Ser Val Val Arg Gly Gly Leu Val Arg Val Gly Gly Asn
195 200 205

Gly Pro Gly Ser Ile Ser Ser Phe Asp Leu Pro Glu Asp His Asp Leu
210 215 220

Val Ala Gly Arg Gly Trp Tyr Trp Val Leu His Gly Arg Pro Val Ser
225 230 235 240

Gly Thr Gly Glu Cys Val Ser Ala Val Ala Ala Ala His Gly Arg Asp
245 250 255

Ala Ala Leu Trp Asp Ala Phe Ala Val Cys Thr Arg Ala Leu Glu Ala
260 265 270

Ala Leu Leu Ser Gly Gly Ser Pro Asp Ala Ala Ile Thr Glu Asn Gln
275 280 285

Arg Leu Leu Glu Arg Ile Gly Val Val Pro Ala Ala Thr Gln Ala Leu
290 295 300

Val Ala Gln Ile Glu Glu Ala Gly Gly Ala Ala Lys Ile Cys Gly Ala
305 310 315 320

Gly Ser Val Arg Gly Asp His Gly Gly Ala Val Leu Val Arg Ile Asp
325 330 335

Asp Ala Gln Ala Met Ala Ser Val Met Ala Arg His Pro Asp Leu Asp
340 345 350

Trp Ala Pro Leu Arg Met Ser Arg Thr Gly Ala Ala Pro Gly Pro Ala
355 360 365

Pro Arg Ala Gln Pro Leu Pro Gly Gln Gly
370 375

<210> 50

<211> 9066

<212> DNA

<213> Paracoccus sp. R114

<220>

<221> CDS

<222> (6970)..(7887)

<223> pmk

<400> 50	
ggatccggca gctcgacacg ccgcagaacc tgtacgaacg tcccgccagc cgttcgtcg	60
cggaattcgt cgggcgcggg acggtgtgc ccgtgcaggg ccatgacggc gcggggccgcg	120
cccgcatcct gggggccgag gtggcggtga acgcccggcc gcaatcgccg tttgtcgatc	180
acgtctgcct gcgcggccgag aaccttgcca tctccgagac gggcgacctg cgcgcacaagg	240
tcgcgcgcgt cacctatctt ggccggaaat acctgctgga aaccgtgctg gattgcggca	300
cccggtcggt gaccgagacc cgccgcggct tcgatacggg cgccgcagctt ggcctgacca	360
tcaacgcggcc ctgggccttt gccgaggatt gaatggacag cgtgaagatc ctttcggca	420
tgggcgtgaa gggccctgcc tcgatcaggc tggatgtcg cgggatgcgc ctgatcctcg	480
attgcgggac cgccgcggac gagggcgccg agttcgaccc cgccctggctg gcggacgcgg	540
atgcgggtgct gatcacccat gaccacgtgg accatatcgg cggcgccgcgt cacgcggctg	600
cggcggggct gccgatccat gcgacgcggc agacggcgaa gttgctgccc gcggggggcg	660
atctgcgcct gctgcccga cgcgggtgtca cgcggatcgc cggggtcgat ctgacgaccg	720
gtcgcaacgg gcatgcccgcg ggccgcgtct ggatgcattt cgacatgggc gaggggctgt	780
tctattccgg cgactggtcc gaggaatccg actggttcgc cttcgatccg ccccccgcctg	840
cggggacggc gattctcgac tgctcctatg gcggtttcga cgtggcgcaa tcggattgca	900
tcgcggaccc ggacgacccgt ctcgagggtgc tgccggggca ggtactgctg cccggccgc	960
catccggccg cgccgcggag ctggccctgc ggctgatccg cccgcacggc ccgggcagcg	1020
tgtatggtcga cgacgcctgc ctgcggcca tcgcgcaact ccccgaggcg cgcggactgg	1080
cctacgccac cgaggcacgc ttcttgcgtc gcgacacgc gAACGCCGAA AGCCGGCGCG	1140
gcatggcggc atctgcaagc atggcgatgc ggggcaggc tggggcgaaa cgcgcatttc	1200
gtcttcaccg ggcacatgaa cgtccatgcg cgcgcatttc gcgaccgc cggcgggcatt	1260
ttccgcgcgt ggaacgtgca tccgcgcgtc cgcgaccaggc gacggatgct ggaacggctg	1320
gcgcgcggc gctttccccc ggccttctgc cccgaccccg agatctatct ggcgcgtggac	1380

atgggcgcgc	aggctttcat	gcaccaggag	gtgacgccat	gatccccgcc	cgcagttct	1440
gcctgatccg	ccacggcgaa	acgaccgcca	atgcaggggc	gatcatcgcg	ggcgcaaccg	1500
atgtccccct	gacgccaagg	ggccgcgatc	aggcccgcgc	cctggcaggg	cgcgaatggc	1560
catcgggcat	cgcgctgttc	gccagcccga	tgtcgctgc	ccgcgataacc	gcgctgctgg	1620
cctttccggg	gcgcgaccac	cagcccgaac	ccgatctgcg	cgaacgcgac	tggggcatct	1680
tcgagggacg	ccccgtcgcc	gatctgcccc	cgcgcgaaat	cacgcccgcag	gggggcgagg	1740
gctgggacga	cgtgatggcc	cgcgtggacc	gcgcgatccg	gcggatctgc	gcgacacctcg	1800
gcgatgcgct	gccggtgctg	gtctgccatt	cggcgtgtat	ccgtgccgcg	cgcgtgctgt	1860
ggaccaccgg	cgtgcgggc	gatgtccgc	ccaacgcac	gccgatcctg	ttcagcccg	1920
acggcgaccg	attaaaggaa	ggaacgatat	gaccgccacc	acccccgtcg	tcgtcttcga	1980
acgtggacgg	cacgcttgcc	gaattcgacg	ccgaccgcct	ggccatctt	gtccacggca	2040
cgaccaagca	ctgggacgcc	ttccaccacg	cgatggccga	cgcggccccc	atccccgagg	2100
tcgccccct	gatgcgcaag	ctgaaggagg	ggggcgagac	ggtcgtcatac	tgctcggggc	2160
ggccccgcgg	ctggcaggat	cagacgatcg	catggctgcg	caagcacgac	ctgcccattcg	2220
acgggatcta	tctgcgcccc	gaggatcagg	acggcgccag	cgaccccgag	gtcaagcgcc	2280
gcgccttagc	cgagatgcgc	gccgacgggc	tggcgcctg	gctggtcgtg	gacgaccggc	2340
ggtcggcgt	ggatgcctgg	cggccgagg	ggctggtctg	cctgcaatgc	gcgcgggggg	2400
acttctaggg	ccgcgcgacg	ggggcgcgga	caggctggc	ggaaaaccgc	ccgcacacca	2460
tgtcctgcac	gcgtcgaaacc	gccccgtccga	cgcgggttcc	cgcacggaaa	cgcgcggcaa	2520
gttgacataa	cttgcacgca	acgtctcgat	tctgcccgcg	aagaatgcga	tgcacatccaga	2580
tgatgcagaa	cgaagaagcg	gaagcgcccc	tgaaagacca	gatgatttcc	catacccg	2640
tgcccacgca	atgggtcgcc	ccgatcctgt	tccgcggccc	cgtcgctcgag	ggcccgatca	2700
gcgcgcccgt	ggccacctac	gagacgcccgc	tctggccctc	gaccgcgcgg	ggggcaggggg	2760
tttccccggca	ttcgggcggg	atccaggtct	cgctggtcga	cgaacgcatg	agccgctcga	2820
tcgcgctgcg	ggcgcgtgac	ggggcggcg	cgaccgcgc	ctggcagtgc	atcaaggccc	2880
gccaggaaga	ggtcgcggcc	gtggtcgcca	ccaccagccg	cttcgcggcgc	cttgcgagc	2940
tgaatcgcca	gatcggtggc	aacctgcttt	acatccgcac	cgaatgcgtg	acgggcgacg	3000
cctcgggtca	caacatggtc	accaaggccg	ccgaggccgt	gcagggctgg	atcctgtcgg	3060

aatacccgat	gctggcctat	tccacgatct	cggggAACCT	gtgcaccgac	aagaaggcgt	3120
cgcggtcaa	cggcatcctg	ggccgcggca	aatacgcgt	cgcgcaggtc	gagatcccgc	3180
gcaagatcct	gacccgcgtg	ctgcgcacca	gcgcgcagaa	gatggtccgc	ctgaactacg	3240
agaagaacta	tgtcggggg	acgctggcgg	ggtcgctgcg	cagtgcgaac	gcgcatttcg	3300
ccaacatgct	gctgggcttc	tacctggcga	cggggcagga	cgcggccaac	atcatcgagg	3360
ccagccaggg	cttcgtccat	tgcgaggccc	gcggcgagga	tctgtatttc	tcgtgcacgc	3420
tgcccaacct	catcatgggc	tcggtcggtg	ccggcaaggg	catccccctcg	atcgaggaga	3480
acctgtcgcg	gatgggctgc	cgcgcgggg	gcgaaccgg	cgacaacgcg	cgccgtcttg	3540
cggcgatctg	cgcggcgtc	gtgctgtgtg	gtgaattgtc	gctgcttgcg	gcccgaccca	3600
accccgagaa	gttggtccgc	acccacatgg	agatggagcg	atgaccgaca	gcaaggatca	3660
ccatgtcgcg	ggcgcaagc	tggaccatct	gcgtgcattt	gacgacgatg	cgatatcgaa	3720
ccggggcgac	agcggcttcg	accgcattcgc	gctgaccat	cgcgccttc	ccgagggtgga	3780
tttcgacgcc	atcgacacgg	cgaccagctt	cctggccgt	gaactgtcct	tcccgctgct	3840
gatctcgcc	atgaccggcg	gcaccggcga	ggagatcgag	cgcataacc	gcaacctggc	3900
cgcgtgtgcc	gaggaggccc	gcgtgcatt	ggcggtggc	tcgcagcgcg	tgatgttcac	3960
cgaccctcg	gcbcgggcca	gcttcgacct	gcgcgcatt	gcgcacc	tgccgctgct	4020
ggccaatatc	ggcgcggtgc	agctgaacat	ggggctgggg	ctgaaggaat	gcctggccgc	4080
gatcgaggtg	ctgcaggcgg	acggcctgta	tctgcacctg	aacccctcg	aagaggccgt	4140
ccagcccgag	ggggatcgcg	actttgcga	tctggcagc	aagatcgcg	ccatcgcccc	4200
cgacgttccc	gtgcccgtcc	tgctgaagga	ggtgggctgc	ggctgtcgg	cgcccgat	4260
cgccatcggg	ctgcgcgcgg	ggatccggca	tttcgacgtg	gccggcgcgc	gcggcacatc	4320
ctggagccgg	atcgagtatc	gccgcggcca	gcggggccat	gacgacctgg	gcctggtctt	4380
ccaggactgg	ggcctgcaga	ccgtggacgc	cctgcgcgag	gcgcggcccg	cgcttgcggc	4440
ccatgatgga	accagcgtgc	tgatcgccag	cggcgccatc	cgcaacggtg	tcgacatggc	4500
gaaatgcgtc	atccctgggg	ccgacatgtg	cggggctgcc	gcgcacc	tgaaagcgac	4560
ccaaaactcg	cgcgaggcgg	ttgtatccgc	catccggaaa	ctgcacatgg	agttccggac	4620
agccatgttc	ctccctgggtt	gcggcacgt	tgccgacctg	aaggacaatt	cctcgctt	4680
ccgtcaatga	aagtgcctaa	gatgaccgtg	acaggaatcg	aagcgatcag	cttctacacc	4740
ccccagaact	acgtggact	ggatatcctt	gccgcgcattc	acgggatcga	ccccgagaag	4800

ttctcgagg ggatcgggca ggagaaaatc gcactgccc gccatgacga ggatatcgta	4860
accatggccg ccgaggccgc gctgccgatc atcgaacgcg cggcacgca gggcatcgac	4920
acggttctgt tcgcccaccga gagcgggatc gaccagtcta agggccgcgc catctatctg	4980
cggccctgc tggacctgtc gcccaaactgc cggtgcgtcg agctgaagca ggctgctat	5040
tccgcgacgg cggcgctgca gatggccctgc gcgcattgtcg cccgcaagcc cgaccgcaag	5100
gtgctggta tcgcgtccga tgtcgcgcgc tatgaccgcg aaagctcgaa cgaggcgacg	5160
cagggtgcgg ggcgcgtcgc catccttgc acgcggatc ccaaggtggc cgagatcgcc	5220
accgtctcgg ggctgttcac cgaggatatac atggatttct ggccggccaa ccaccgcccgc	5280
acgccccctgt tcgacggcaa ggcattcgacg ctgcgtatc tgaacgcgt ggtcgaggcg	5340
tggaacgact atcgcgcgaa tggcggccac gagttcgccg atttcgcgca tttctgctat	5400
cacgtccgt tctcgccggat gggcgagaag gcgaacagcc acctggccaa ggcgaacaag	5460
acgcccgtgg acatggggca ggtcagacg ggcctgatct acaaccggca ggtcggaaac	5520
tgctataccg ggtcgatcta cctggcattc gcctcgctgc tggagaacgc tcaggaggac	5580
ctgaccggcg cgctggtcgg tctgttcagc tatggctcgg gtgcgacggg cgaattttc	5640
gatgcgcgga tcgcgcggg ttaccgcgac cacctgtcg cgaaacgcca tcgcgaatttgc	5700
ctgcaggatc gcacgcccgt cacatatgac gaatacgttgc ccctgtggga cgagatcgac	5760
ctgacgcagg ggcgcggcga caaggcgccg ggtcgttca ggctggcagg tatcgaggac	5820
gagaagcgca tctatgtcga ccggcaggcc tgaagcaggc gccatgccc cggcaagct	5880
gatcctgtcc gggaaacatt ccgtgtctta tggcgcccc ggcattgcca tggccatcg	5940
ccgctataacc gaggtgtggt tcacgcccgt tggcattggc gagggatac gcacgacatt	6000
cgccaatctc tcgggggggg cgaccttattc gctgaagctg ctgtcggtt tcaagtgcgc	6060
gctggaccgc cggttcgagc agttcctgaa cggcgaccta aaggtgcaca aggtcctgac	6120
ccatccccac gatctggcggt tctatgcgt ggcgtcgctt ctgcacgaca agccgccccgg	6180
gaccggccgcg atgccgggca tcggcgcgat gcaccacctg cggcgaccgg gtgagctgg	6240
cagccggacg gagctgccc tcggcgccgg catgggtcg tctgcggcca tcgtcgccgc	6300
caccacggtc ctgttcgaga cgctgctgga cggccccaa acgcccgaac agcgcttcga	6360
ccgcgtccgc ttctgcgagc ggttgaagca cggcaaggcc ggtccatcg acgcggccag	6420
cgtcggtcgcc ggccggcttg tccgcgtggg cgggaacggg ccgggttcga tcagcagctt	6480

cgatggccc gaggatcacg accttgcgc gggacgcggc tggtaactggg tactgcacgg	6540
gcgcggccgtc agcgggaccg gcgaatgcgt cagcgcggtc gcggcggcgc atggtcgcga	6600
tgcggcgctg tgggacgcct tcgcagtctg cacccgcgcg ttggaggccg cgctgctgtc	6660
tgggggcagc cccgacgccc ccatcaccga gaaccagcgc ctgctgaaac gcatcggcgt	6720
cgtgccggca gcgacgcagg ccctcggtgc ccagatcgag gaggcgggtg gcgcggccaa	6780
gatctgcggc gcagggttccg tgccccggcga tcacggcggg gcggtcctcg tgcggattga	6840
cgacgcgcag gcgatggctt cggatggcgc ggcgcattccc gacctcgact gggcgccccct	6900
gcgcatgtcg cgacacggggg cggcacccgg ccccgccgcg cgtgcgcaac cgctgcggg	6960
gcagggctg atg gat cag gtc atc cgc gcc agc gcg ccg ggt tcg gtc atg Met Asp Gln Val Ile Arg Ala Ser Ala Pro Gly Ser Val Met	7011
1 5 10	
atc acg ggc gaa cat gcc gtg gtc tat gga cac cgc gcc atc gtc gcc Ile Thr Gly Glu His Ala Val Val Tyr Gly His Arg Ala Ile Val Ala	7059
15 20 25 30	
ggg atc gag cag cgc gcc cat gtg acg atc gtc ccg cgt gcc gac cgc Gly Ile Glu Gln Arg Ala His Val Thr Ile Val Pro Arg Ala Asp Arg	7107
35 40 45	
atg ttt cgc atc acc tcg cag atc ggg gcg ccg cag cag ggg tcg ctg Met Phe Arg Ile Thr Ser Gln Ile Gly Ala Pro Gln Gln Gly Ser Leu	7155
50 55 60	
gac gat ctg cct gcg ggc acc tat cgc ttc gtg ctg gcc gcc atc Asp Asp Leu Pro Ala Gly Gly Thr Tyr Arg Phe Val Leu Ala Ala Ile	7203
65 70 75	
gcg cga cac gcg ccg gac ctg cct tgc ggg ttc gac atg gac atc acc Ala Arg His Ala Pro Asp Leu Pro Cys Gly Phe Asp Met Asp Ile Thr	7251
80 85 90	
tcg ggg atc gat ccg agg ctc ggg ctt gga tcc tgc gcg gcg gtg acg Ser Gly Ile Asp Pro Arg Leu Gly Leu Gly Ser Ser Ala Ala Val Thr	7299
95 100 105 110	
gtc gcc tgc ctc ggc gcg ctg tcg cgg ctg gcg ggg cgg ggg acc gag Val Ala Cys Leu Gly Ala Leu Ser Arg Leu Ala Gly Arg Gly Thr Glu	7347
115 120 125	
ggg ctg cat gac gac gcg ctg cgc atc gtc cgc gcc atc cag ggc agg Gly Leu His Asp Asp Ala Leu Arg Ile Val Arg Ala Ile Gln Gly Arg	7395
130 135 140	
ggc agc ggg gcc gat ctg gcg gcc agc ctg cat ggc ggc ttc gtc gcc Gly Ser Gly Ala Asp Leu Ala Ala Ser Leu His Gly Gly Phe Val Ala	7443
145 150 155	
tat cgc gcg ccc gat ggc ggt gcc gcg cag atc gag gcg ctt ccg gtg	7491

Tyr Arg Ala Pro Asp Gly Gly Ala Ala Gln Ile Glu Ala Leu Pro Val			
160	165	170	
ccg ccg ggg ccg ttc ggc ctg cgc tat gcg ggc tac aag acc ccg aca			7539
Pro Pro Gly Pro Phe Gly Leu Arg Tyr Ala Gly Tyr Lys Thr Pro Thr			
175	180	185	190
gcc gag gtg ctg cgc ctt gtg gcc gat cggt atg ggc aac gag gcc			7587
Ala Glu Val Leu Arg Leu Val Ala Asp Arg Met Ala Gly Asn Glu Ala			
195	200	205	
gct ttc gac gcg ctc tac tcc cgg atg ggc gca agc gca gat gcc gcg			7635
Ala Phe Asp Ala Leu Tyr Ser Arg Met Gly Ala Ser Ala Asp Ala Ala			
210	215	220	
atc cgc gcg caa ggg ctg gac tgg gct gca ttc cac gac gcg ctg			7683
Ile Arg Ala Ala Gln Gly Leu Asp Trp Ala Ala Phe His Asp Ala Leu			
225	230	235	
aac gaa tac cag cgc ctg atg gag cag ctg ggc gtg tcc gac gac acg			7731
Asn Glu Tyr Gln Arg Leu Met Glu Gln Leu Gly Val Ser Asp Asp Thr			
240	245	250	
ctg gac gcg atc atc cgc gag gcg cgc gac gcg ggc gcc gca gtc gcc			7779
Leu Asp Ala Ile Ile Arg Glu Ala Arg Asp Ala Gly Ala Ala Val Ala			
255	260	265	270
aag atc tcc ggc tcg ggg ctg ggg gat tgc gtg ctg gca ctg ggc gac			7827
Lys Ile Ser Gly Ser Gly Leu Gly Asp Cys Val Leu Ala Leu Gly Asp			
275	280	285	
cag ccc aag ggt ttc gtg ccc gca agc att gcc gag aag gga ctt gtt			7875
Gln Pro Lys Gly Phe Val Pro Ala Ser Ile Ala Glu Lys Gly Leu Val			
290	295	300	
ttc gat gac tga tgccgtccgc gacatgatcg cccgtgccat ggccggcg			7927
Phe Asp Asp			
305			
accgacatcc gagcagccga ggcttatgcg cccagcaaca tcgcgtgtc gaaatactgg			7987
ggcaagcgcg acgcccgcgc gaaccttccg ctgaacagct ccgtctcgat ctcgttggcg			8047
aactggggct ctcatacgcg ggtcgagggg tccggcacgg gccacgacga ggtgcacac			8107
aacggcacgc tgctggatcc gggcgacgcc ttgcgcgcgc ggcgttggc attcgctgac			8167
ctgttccggg gggggaggca cctgcccgtg cggatcacga cgcagaactc gatcccacg			8227
gcggcggggc ttgcctcgat ggcctcgaaa ttgcggcgat tgacccgtgc gctggcgaaa			8287
gcgttcgggc tggatctgga cgacacggat ctgagccgca tcgccccggat cggcagtggc			8347
agcgccgccc gctcgatctg gcacggcttc gtccgctgga accggggcgaa ggccgaggat			8407
gggcattgaca gccacggcgat cccgctggac ctgcgtggc ccggcttccg catcgatc			8467

gtggccgtgg acaaggggcc caagccttc agttcgcgcg acggcatgaa ccacacggtc	8527
gagaccagcc cgctgttccc gccctggcct gcgcaggcg aagcggattg ccgcgtcatc	8587
gaggatgcga tcgcccggcc cgacatggcc gccctggtc cgcggtcga ggcgaacgcc	8647
cttgcgatgc acgccacgat gatggccgcg cgcgcgcgc tctgtaccc gacgggcggc	8707
agctggcagg tgctggaacg cctgtggcag gcccgcgcgg acgggcttgc ggccttgcg	8767
acgatggatg cggccccgaa cgtcaagctg atcttcgagg aaagcagcgc cgcgcacgtg	8827
ctgtacctgt tccccgacgc cagcctgatc gcgcgcgtcg aggggcgttg aacgcgtaag	8887
acgaccactg ggttaaggttc tgccgcgcgt ggtctcgact gcctgcaaag aggtgcttga	8947
gttgctgcgt gactgcggcg gcccacttcg tgggacttgc cgcgcacgct gacgcgcgtgg	9007
aaacgcgccc gcgattacg accgcgtcat tgccctgaac caattcccg tcggtcgac	9066

<210> 51

<211> 305

<212> PRT

<213> Paracoccus sp. R114

<400> 51

Met Asp Gln Val Ile Arg Ala Ser Ala Pro Gly Ser Val Met Ile Thr			
1	5	10	15

Gly Glu His Ala Val Val Tyr Gly His Arg Ala Ile Val Ala Gly Ile			
20	25	30	

Glu Gln Arg Ala His Val Thr Ile Val Pro Arg Ala Asp Arg Met Phe			
35	40	45	

Arg Ile Thr Ser Gln Ile Gly Ala Pro Gln Gln Gly Ser Leu Asp Asp			
50	55	60	

Leu Pro Ala Gly Gly Thr Tyr Arg Phe Val Leu Ala Ala Ile Ala Arg			
65	70	75	80

His Ala Pro Asp Leu Pro Cys Gly Phe Asp Met Asp Ile Thr Ser Gly			
85	90	95	

Ile Asp Pro Arg Leu Gly Leu Gly Ser Ser Ala Ala Val Thr Val Ala
100 105 110

Cys Leu Gly Ala Leu Ser Arg Leu Ala Gly Arg Gly Thr Glu Gly Leu
115 120 125

His Asp Asp Ala Leu Arg Ile Val Arg Ala Ile Gln Gly Arg Gly Ser
130 135 140

Gly Ala Asp Leu Ala Ala Ser Leu His Gly Gly Phe Val Ala Tyr Arg
145 150 155 160

Ala Pro Asp Gly Gly Ala Ala Gln Ile Glu Ala Leu Pro Val Pro Pro
165 170 175

Gly Pro Phe Gly Leu Arg Tyr Ala Gly Tyr Lys Thr Pro Thr Ala Glu
180 185 190

Val Leu Arg Leu Val Ala Asp Arg Met Ala Gly Asn Glu Ala Ala Phe
195 200 205

Asp Ala Leu Tyr Ser Arg Met Gly Ala Ser Ala Asp Ala Ala Ile Arg
210 215 220

Ala Ala Gln Gly Leu Asp Trp Ala Ala Phe His Asp Ala Leu Asn Glu
225 230 235 240

Tyr Gln Arg Leu Met Glu Gln Leu Gly Val Ser Asp Asp Thr Leu Asp
245 250 255

Ala Ile Ile Arg Glu Ala Arg Asp Ala Gly Ala Ala Val Ala Lys Ile
260 265 270

Ser Gly Ser Gly Leu Gly Asp Cys Val Leu Ala Leu Gly Asp Gln Pro
275 280 285

Lys Gly Phe Val Pro Ala Ser Ile Ala Glu Lys Gly Leu Val Phe Asp
290 295 300

Asp
305

<210> 52

<211> 9066

<212> DNA

<213> Paracoccus sp. R114

<220>

<221> CDS

<222> (7880)..(8878)

<223> mvd

<400> 52		
ggatccggca gctcgacacg ccgcagaacc tgtacgaacg tcccgccagc cgcttcgtcg	60	
cggaattcgt cgggcgcggg acggtgtgc ccgtgcaggc ccatgacggc gcggggccgcg	120	
cccgcatcct gggggccgag gtggcggta acgcccggcc gcaatcgccgc tttgtcgatc	180	
acgtctgcct gcgcggccgag aaccttgcca tctccgagac gggcgacctg cgccgcggcaagg	240	
tgcgcgcgt cacctatctt ggcggaaat acctgctgga aaccgtgctg gattgcggca	300	
cccggttgtt gaccgagacc cgccgcggct tcgatacggg cgccgacgtt ggccctgacca	360	
tcaacgcccc ctgggccttt gccgaggatt gaatggacag cgtgaagatc ctttcggca	420	
tgggcgtgaa gggccctgcc tcgatcaggc tggatgtcgg cggatgcgc ctgatcctcg	480	
attgcgggac cggccggac gagggcgcgg agttcgaccc cgccctggctg gccggacgcgg	540	
atgcgggtct gatcacccat gaccacgtgg accatatcgg cggcgccgt cacgcggctcg	600	
cggcgggct gccgatccat gcgacgcggc agacggcggg gttgtgccc gcggggggcgg	660	
atctgcgcct gctgcccggaa cgcgggtgtca cgcggatcgc cggggatcgat ctgacgaccg	720	
gtcgcaacgg gcatgccgcg ggcggcgtct ggatgcattt cgacatggc gaggggctgt	780	
tctattccgg cgactggtcc gaggaatccg actgggttcgc cttcgatccg ccccccgcctg	840	
cggggacggc gattctcgac tgctccatg gcgggttcga cgtggcgcaa tcggattgca	900	
tcgcggaccc ggacgacactg ctcgaggtgc tgccggggca ggtactgctg cccggatccgc	960	
catccggccg cgcggccgag ctggccctgc ggctgatccg cgcgcacggc cccggcagcg	1020	
tgatggtcga cgacgcctgc ctgcccggcca tcgcgcaact gcccggaggcg cgcggactgg	1080	
cctacgcac cggggcacgc ttcttgcgtc gcgcacacgca gacgcggcaaa agccggcgcg	1140	

gcatggcggc atctgcaagc atggcgcat gcgggcaggc tggggcggga cgcgcattgtc 1200
 gtcttcaccg ggcacatgaa cgtccatgcg cgccattct gcgaccgccc cggcgggcat 1260
 ttccgcccgtt ggaacgtgca tccgccgtg cgccaccaggc gacggatgct ggaacggctg 1320
 gccgcgcggc gctttccccc ggccattctgc cccgaccccg agatctatct ggctggac 1380
 atgggcgcgc aggtcttcat gcaccaggag gtgacgccat gatccccgcc cgcaagttct 1440
 gcctgatccg ccacggcgaa acgaccgcca atgcaggggc gatcatcgcg ggcaaccg 1500
 atgtgcccct gacgccaagg ggccgcgatc aggccgcgc cctggcaggg cgcaatggc 1560
 catcgggcat cgcgctgttc gccagccga tgtcgcgtgc ccgcataacc ggcgtgtgg 1620
 cctttccggg ggcgcgaccac cagccgaac ccgatctgcg cgaacgcgac tgggcattct 1680
 tcgagggacg cccgtcgcc gatctgcccc cgccgaaat cacgcccag gggggcgagg 1740
 gctgggacga cgtgatggcc cgctggacc ggcgcattccg gccgatctgc ggcacctcgg 1800
 gcgatgcgtt gccggtgctg gtctgccatt cggcggtgat ccgtgccgcg cgctgtgt 1860
 ggaccaccgg cgatgcgggc gatgtccgc ccaacgcac gccgatcctg ttccgggg 1920
 acggcgaccg attaaaggaa ggaacgatat gaccgccacc accccctgcg tcgtttcga 1980
 acgtggacgg cacgcttgcc gaattcgacg ccgaccgcct gggccatctt gtccacggca 2040
 cgaccaagca ctgggacgcc ttccaccacg cgatggccga cgcccccggc atccccgagg 2100
 tcgccccct gatgcgaag ctgaaggagg gggcgagac ggtcgatcgc tgctcgggc 2160
 ggccccggg ctggcaggat cagacgatcg catggctgcg caagcacgac ctgccttcg 2220
 acgggatcta tctgcgcacc gaggatcagg acggcgccag cgaccccgag gtcaagcgcc 2280
 ggcgccttagc cgagatgcgc gccgacgggc tggcgccctg gctggcgatc gacgaccggc 2340
 ggtccgtcgt ggatgcctgg cggcccgagg ggctggctcg cctgcaatgc ggcgggggg 2400
 acttcttaggg ccgcgcgacg gggcgccgga caggctggc gggaaaccgc cccgcacca 2460
 tgtcctgcac gcgtcgaaacc gcccgtccga cgccggtttc cgacggaaa cgccggcaa 2520
 gttgacataa cttgcacgcg acgtctcgat tctgcccgcg aagaatgcg a tgcacccaga 2580
 tgatgcagaa cgaagaagcg gaagcgccc tgaaagacca gatgattcc catacccg 2640
 tgcccacgca atgggtcgcc ccgatctgt tccgcggccc cgctcgatc ggcggatca 2700
 gcgccgcgtt ggccacctac gagacgccgc tctggccctc gaccgcgcgg gggcgagg 2760
 ttccggca ttccggcgaa atccaggtct cgctggtcga cgaacgcattg agccgctcga 2820

tcgcgcgtcg	ggcgcatgac	ggggcggcgg	cgaccgcccgc	ctggcagtcg	atcaaggccc	2880
gccaggaaga	gttcgcggcc	gtggtcgcca	ccaccagccg	cttcgcccgc	cttgtcgagc	2940
tgaatcgcca	gatcgtggc	aacctgctt	acatccgcat	cgaatgcgtg	acggggcgacg	3000
cctcgggtca	caacatggtc	accaaggccg	ccgaggccgt	gcagggctgg	atcctgtcgg	3060
aatacccgat	gctggcctat	tccacgatct	cggggAACCT	gtgcaccgac	aagaaggcgt	3120
cggcggtcaa	cggcattcctg	ggccgcggca	aatacgccgt	cgccgaggtc	gagatcccgc	3180
gcaagatcct	gacccgcgtg	ctgcgcacca	gcccgcgagaa	gatggtccgc	ctgaactacg	3240
agaagaacta	tgtcgggggt	acgctggcgg	ggtcgctcg	cagtgcgaac	gcgcatttcg	3300
ccaacatgct	gctgggcttc	tacctggcga	cggggcagga	cgccgccaac	atcatcgagg	3360
ccagccaggg	cttcgtccat	tgcgaggccc	gcggcgagga	tctgtatttc	tcgtgcacgc	3420
tgcccaacct	catcatggc	tcggcggtg	ccggcaaggg	catcccctcg	atcgaggaga	3480
acctgtcgcg	gatgggctgc	cggcagccgg	gcgaacccgg	cgacaacgcg	cgccgtcttg	3540
cggcgatctg	cgcggcgtc	gtgctgtgtg	gtgaattgtc	gctgctgcg	gcccagacca	3600
accccggaga	gttggtccgc	acccacatgg	agatggagcg	atgaccgaca	gcaaggatca	3660
ccatgtcgcg	ggcgcaagc	tggaccatct	gcgtgcattg	gacgacgatg	cgatatcga	3720
ccggggcgac	agcggcttcg	accgcacatgc	gctgacccat	cgcccccgtc	ccgaggtgga	3780
tttcgacgcc	atcgacacgg	cgaccagctt	cctggccgt	gaactgtcct	tcccgtgct	3840
gatctcgcc	atgaccggcg	gcacccggcga	ggagatcgag	cgcatcaacc	gcaacctggc	3900
cgctggtgcc	gaggaggccc	gcgtcgccat	ggcgggtggc	tcgcagcgcg	tgtgttcac	3960
cgacccctcg	gcgcgggcca	gcttcgaccc	gcgcgcaccc	tgccgctgct		4020
ggccaatatac	ggcgcggtgc	agctgaacat	ggggctgggg	ctgaaggaat	gcctggccgc	4080
gatcgaggtg	ctgcaggcgg	acggcctgta	tctgcacctg	aacccctgc	aagaggccgt	4140
ccagccccag	ggggatcgcg	actttgccga	tctggcagc	aagatcgccg	ccatcgcccc	4200
cgacgttccc	gtgcccgtcc	tgctgaagga	ggtgggctgc	ggcctgtcgg	cgcccgatat	4260
cgccatcggg	ctgcgcgcgc	ggatccggca	tttcgacgtg	gccggcgcg	gcggcacatc	4320
ctggagccgg	atcgagtatc	gccgcccaca	gcgggcccgt	gacgacctgg	gcctggtctt	4380
ccaggactgg	ggcctgcaga	ccgtggacgc	cctgcgcgag	gcgcggcccg	cgcttgcggc	4440
ccatgatgga	accagcgtgc	tgatcgccag	cgccggcatc	cgcaacggtg	tcgacatggc	4500
gaaatgcgtc	atcctggggg	ccgacatgtg	cggggtcgcc	gcgccttcgc	tgaaagcgcc	4560

ccaaaaactcg cgcgaggcgg ttgtatccgc catccggaaa ctgcacatctgg agttccggac 4620
 agccatgttc ctccctgggtt gcggcacgct tgccgacctg aaggacaatt cctcgcttat 4680
 ccgtcaatga aagtgcctaa gatgaccgtg acaggaatcg aagcgatcg attctacacc 4740
 ccccagaact acgtgggact ggatatcctt gccgcgcatac acgggatcga ccccgagaag 4800
 ttctcgaagg ggatcgggca ggagaaaatc gcactgcccgc gccatgacga ggatatcgtg 4860
 accatggccg ccgaggccgc gctgccgatc atcgaacgcg cgggcacgca gggcatcgac 4920
 acggttctgt tcgcccaccga gagcgggatc gaccagtca gaggccgcgc catctatctg 4980
 cgccgcctgc tggacctgtc gcccaactgc cggtgcgtcg agctgaagca ggcctgctat 5040
 tccgcgacgg cgccgcgtgca gatggcctgc ggcgcgtcg cccgcaagcc cgaccgcaag 5100
 gtgctggta tcgcgtccga tgtcgccgc tatgaccgcg aaagctcggg cgaggcgacg 5160
 cagggtcgaa ggcgcgtcgcatccttgcg acgcgcgatc ccaaggtggc cgagatcgcc 5220
 accgtctcgg ggctgttcac cgaggatatc atggatttct ggcggccgaa ccaccgcgc 5280
 acgcccctgt tcgacggcaa ggcgcgtacg ctgcgcgtatc tgaacgcgt ggtcgaggcg 5340
 tggAACGACT atcgcgcgaa tggcgccac gagttcgccg atttcgcgca tttctgctat 5400
 cacgtgcgt tctcgccgat gggcgagaag ggcgcgtacg acctggccaa ggcgaacaag 5460
 acgcccgtgg acatggggca ggtgcagacg ggcctgatct acaaccggca ggtcgaaaaac 5520
 tgctataaccg ggtcgatcta cctggcattc gcctcgctgc tggagaacgc tcaggaggac 5580
 ctgaccggcg cgctggtcgg tctgttcagc tatggctcgg gtgcgcacggg cgaattcttc 5640
 gatgcgcgga tcgcgcggg ttaccgcgac cacctgttcg cggAACGCCA tcgcgaattg 5700
 ctgcaggatc gcacgcccgt cacatatgac gaatacggttgc cctgtggga cgagatcgac 5760
 ctgacgcagg gcgcgcggcga caaggcgacg ggtcggttca ggctggcagg tatcgaggac 5820
 gagaagcgca tctatgtcga ccggcaggcc tgaagcaggc gcccattgc cggcaagct 5880
 gatcctgtcc gggaaacatt ccgtgatcta tggtgccccc ggcattgc tggccatcg 5940
 ccgctataacc gaggtgtggt tcacgcccgt tggcattggc gagggatac gcacgacatt 6000
 cgccaatctc tcgggcgggg cgacatttc gctgaagctg ctgtcggtt tcaagtcgc 6060
 gctggaccgc cggttcgagc agttcctgaa cggcgaccta aaggtgcaca aggtcctgac 6120
 ccatccgac gatctggcgg tctatgcgt ggcgtcgctt ctgcacgaca agccgcggg 6180
 gaccgcgcg atgcccggca tcggcgat gcaccacctg cggcgaccgg gtgagctggg 6240

cagccggacg gagctgccca tcggcgcccc catggggtcg tctgcggcca tcgtcgccgc 6300
 caccacggtc ctgttcgaga cgctgcttggc ccggcccaag acgcccgaac agcgcttcga 6360
 cccgcgtccgc ttctgcgagc ggttgaagca cggcaaggcc ggtcccatcg acgcggccag 6420
 cgtcgtgcgc ggcgggcttg tccgcgtggc cgggaacggg ccggggttcga tcagcagctt 6480
 cgatttgccc gaggatcacg accttgcgc gggacgcggc tggtaactggg tactgcacgg 6540
 gcgc(cccg)tc agcgggaccg gcgaatgcgt cagcgcggtc gcggcggcgc atggtcgcga 6600
 tgccggcgctg tgggacgcct tcgcagtctg caccgcgcg ttggaggccg cgctgctgtc 6660
 tgggggcagc cccgacgcgg ccatcaccga gaaccagcgc ctgctggAAC gcacgcgt 6720
 cgtgcggca gcgacgcagg ccctcggtgc ccagatcgag gaggcgggtg gcgcggccaa 6780
 gatctgcggc gcaggttccg tgccgggcga tcacggcggt gcggtcctcg tgccgattga 6840
 cgacgcgcag gcgatggctt cggcatggc gcgcacatccc gacctcgact gggcgc(cc)ct 6900
 gcgcacatgcg cgcacggggg cggcacccgg cccgcgcgg cgtgcgaac cgctgcgggg 6960
 gcagggctga tggatcaggt catccgcgc acgcgcggg gttcggtcat gatcacggc 7020
 gaacatgccg tggtaatgg acaccgcgc atcgtcgccg ggatcgagca gcgcgc(cc)at 7080
 gtgacgatcg tcccgcgtgc cgaccgcatg ttcgcatca cctcgcatat cggggcgcgg 7140
 cagcaggggt cgctggacga tctgcctgcg ggccggaccc atcgcttcgt gctggccgc 7200
 atcgcgccgac acgcgcggg cctgccttgc ggttcgaca tggacatcac ctcggggatc 7260
 gatccgagggc tcgggcttgg atcctcgccg gcgggtgacgg tcgcctgcct cggcgcgtc 7320
 tcgcggctgg cggggcgggg gaccgagggg ctgcatgacg acgcgcgtgc catcgccgc 7380
 gccatccagg gcaggggcag cggggccgat ctggcggcca gcctgcattgg cggcttcgtc 7440
 gcctatcgccg cgcgcgtgg cgggtgcgcg cagatcgagg cgcttcgtt ggcgcgggg 7500
 ccgttcggcc tgcgctatgc gggctacaag accccgacag ccgaggtgct gcgccttgc 7560
 gccgatcgga tggcggccaa cgagggcgct ttcgacgcgc tctactcccg gatggcgca 7620
 agcgcgatcg ccgcgatccg cgcggcgcaa gggctggact gggctgcatt ccacgacgcg 7680
 ctgaacgaat accagcgccct gatggagcag ctggcggtgt ccgacgcacac gctggacgcg 7740
 atcatccgcg aggcgccgca cgcggcgcc gcagtcgcca agatctccgg ctcggggctg 7800
 ggggattgcg tgctggact gggcgaccag cccaaagggtt tcgtgc(cc)gc aagcattgcc 7860
 gagaaggac ttgtttcg atg act gat gcc gtc cgc gac atg atc gcc cgt 7912
 Met Thr Asp Ala Val Arg Asp Met Ile Ala Arg
 1 5 10

gcc atg gcg ggc gcg acc gac atc cga gca gcc gag gct tat gcg ccc Ala Met Ala Gly Ala Thr Asp Ile Arg Ala Ala Glu Ala Tyr Ala Pro 15	20	25	7960
agc aac atc gcg ctg tcg aaa tac tgg ggc aag cgc gac gcc gcg cg Ser Asn Ile Ala Leu Ser Lys Tyr Trp Gly Lys Arg Asp Ala Ala Arg 30	35	40	8008
aac ctt ccg ctg aac agc tcc gtc tcg atc tcg ttg gcg aac tgg ggc Asn Leu Pro Leu Asn Ser Ser Val Ser Ile Ser Leu Ala Asn Trp Gly 45	50	55	8056
tct cat acg cgg gtc gag ggg tcc ggc acg ggc cac gac gag gtg cat Ser His Thr Arg Val Glu Gly Ser Gly Thr Gly His Asp Glu Val His 60	65	70	8104
cac aac ggc acg ctg ctg gat ccg ggc gac gcc ttc gcg cgc cgc gcg His Asn Gly Thr Leu Leu Asp Pro Gly Asp Ala Phe Ala Arg Arg Ala 80	85	90	8152
ttg gca ttc gct gac ctg ttc cgg ggg agg cac ctg cgg ctg cgg Leu Ala Phe Ala Asp Leu Phe Arg Gly Arg His Leu Pro Leu Arg 95	100	105	8200
atc acg acg cag aac tcg atc ccg acg gcg ggg ctt gcc tcg tcg Ile Thr Thr Gln Asn Ser Ile Pro Thr Ala Ala Gly Leu Ala Ser Ser 110	115	120	8248
gcc tcg ggg ttc gcg gcg ctg acc cgt gcg ctg gcg ggg gcg ttc ggg Ala Ser Gly Phe Ala Ala Leu Thr Arg Ala Leu Ala Gly Ala Phe Gly 125	130	135	8296
ctg gat ctg gac gac acg gat ctg agc cgc atc gcc cgg atc ggc agt Leu Asp Leu Asp Asp Thr Asp Leu Ser Arg Ile Ala Arg Ile Gly Ser 140	145	150	8344
ggc agc gcc gcc cgc tcg atc tgg cac ggc ttc gtc cgc tgg aac cgg Gly Ser Ala Ala Arg Ser Ile Trp His Gly Phe Val Arg Trp Asn Arg 160	165	170	8392
ggc gag gcc gag gat ggg cat gac agc cac ggc gtc cgc ctg gac ctg Gly Glu Ala Glu Asp Gly His Asp Ser His Gly Val Pro Leu Asp Leu 175	180	185	8440
cgc tgg ccc ggc ttc cgc atc gcg atc gtg gcc gtg gac aag ggg ccc Arg Trp Pro Gly Phe Arg Ile Ala Ile Val Ala Val Asp Lys Gly Pro 190	195	200	8488
aag cct ttc agt tcg cgc gac ggc atg aac cac acg gtc gag acc agc Lys Pro Phe Ser Ser Arg Asp Gly Met Asn His Thr Val Glu Thr Ser 205	210	215	8536
ccg ctg ttc ccg ccc tgg cct gcg cag gcg gaa gcg gat tgc cgc gtc Pro Leu Phe Pro Pro Trp Pro Ala Gln Ala Glu Ala Asp Cys Arg Val 220	225	230	8584

atc gag gat gcg atc gcc gcc cgc gac atg gcc gcc ctg ggt ccg cgg Ile Glu Asp Ala Ile Ala Ala Arg Asp Met Ala Ala Leu Gly Pro Arg 240 245 250	8632
gtc gag gcg aac gcc ctt gcg atg cac gcc acg atg atg gcc gcg cgc Val Glu Ala Asn Ala Leu Ala Met His Ala Thr Met Met Ala Ala Arg 255 260 265	8680
ccg ccg ctc tgc tac ctg acg ggc ggc agc tgg cag gtg ctg gaa cgc Pro Pro Leu Cys Tyr Leu Thr Gly Gly Ser Trp Gln Val Leu Glu Arg 270 275 280	8728
ctg tgg cag gcc cgc gcg gac ggg ctt gcg gcc ttt gcg acg atg gat Leu Trp Gln Ala Arg Ala Asp Gly Leu Ala Ala Phe Ala Thr Met Asp 285 290 295	8776
gcc ggc ccg aac gtc aag ctg atc ttc gag gaa agc agc gcc gcc gac Ala Gly Pro Asn Val Lys Leu Ile Phe Glu Ser Ser Ala Ala Asp 300 305 310 315	8824
gtg ctg tac ctg ttc ccc gac gcc agc ctg atc gcg ccg ttc gag ggg Val Leu Tyr Leu Phe Pro Asp Ala Ser Leu Ile Ala Pro Phe Glu Gly 320 325 330	8872
cgt tga acgcgtaaga cgaccactgg gtaaggttct gccgcgcgtg gtctcgactg Arg	8928
cctgcaaaga ggtgcttgag ttgctgcgtg actgcggcgg ccgacttcgt gggacttgcc	8988
cgccacgctg acgcgctgga aacgcgccccg cggattacga ccgcgtcatt gccctgaacc	9048
aatttcccgtaatcgac	9066

<210> 53

<211> 332

<212> PRT

<213> Paracoccus sp. R114

<400> 53

Met Thr Asp Ala Val Arg Asp Met Ile Ala Arg Ala Met Ala Gly Ala 1 5 10 15
--

Thr Asp Ile Arg Ala Ala Glu Ala Tyr Ala Pro Ser Asn Ile Ala Leu 20 25 30

Ser Lys Tyr Trp Gly Lys Arg Asp Ala Ala Arg Asn Leu Pro Leu Asn 35 40 45

Ser Ser Val Ser Ile Ser Leu Ala Asn Trp Gly Ser His Thr Arg Val
50 55 60

Glu Gly Ser Gly Thr Gly His Asp Glu Val His His Asn Gly Thr Leu
65 70 75 80

Leu Asp Pro Gly Asp Ala Phe Ala Arg Arg Ala Leu Ala Phe Ala Asp
85 90 95

Leu Phe Arg Gly Gly Arg His Leu Pro Leu Arg Ile Thr Thr Gln Asn
100 105 110

Ser Ile Pro Thr Ala Ala Gly Leu Ala Ser Ser Ala Ser Gly Phe Ala
115 120 125

Ala Leu Thr Arg Ala Leu Ala Gly Ala Phe Gly Leu Asp Leu Asp Asp
130 135 140

Thr Asp Leu Ser Arg Ile Ala Arg Ile Gly Ser Gly Ser Ala Ala Arg
145 150 155 160

Ser Ile Trp His Gly Phe Val Arg Trp Asn Arg Gly Glu Ala Glu Asp
165 170 175

Gly His Asp Ser His Gly Val Pro Leu Asp Leu Arg Trp Pro Gly Phe
180 185 190

Arg Ile Ala Ile Val Ala Val Asp Lys Gly Pro Lys Pro Phe Ser Ser
195 200 205

Arg Asp Gly Met Asn His Thr Val Glu Thr Ser Pro Leu Phe Pro Pro
210 215 220

Trp Pro Ala Gln Ala Glu Ala Asp Cys Arg Val Ile Glu Asp Ala Ile
225 230 235 240

Ala Ala Arg Asp Met Ala Ala Leu Gly Pro Arg Val Glu Ala Asn Ala
245 250 255

Leu Ala Met His Ala Thr Met Met Ala Ala Arg Pro Pro Leu Cys Tyr
260 265 270

Leu Thr Gly Gly Ser Trp Gln Val Leu Glu Arg Leu Trp Gln Ala Arg
275 280 285

Ala Asp Gly Leu Ala Ala Phe Ala Thr Met Asp Ala Gly Pro Asn Val
290 295 300

Lys Leu Ile Phe Glu Glu Ser Ser Ala Ala Asp Val Leu Tyr Leu Phe
305 310 315 320

Pro Asp Ala Ser Leu Ile Ala Pro Phe Glu Gly Arg
325 330

<210> 54

<211> 353

<212> PRT

<213> Streptomyces sp. strain CL190

<400> 54

Met Thr Glu Thr His Ala Ile Ala Gly Val Pro Met Arg Trp Val Gly
1 5 10 15

Pro Leu Arg Ile Ser Gly Asn Val Ala Glu Thr Glu Thr Gln Val Pro
20 25 30

Leu Ala Thr Tyr Glu Ser Pro Leu Trp Pro Ser Val Gly Arg Gly Ala
35 40 45

Lys Val Ser Arg Leu Thr Glu Lys Gly Ile Val Ala Thr Leu Val Asp
50 55 60

Glu Arg Met Thr Arg Ser Val Ile Val Glu Ala Thr Asp Ala Gln Thr
65 70 75 80

Ala Tyr Met Ala Ala Gln Thr Ile His Ala Arg Ile Asp Glu Leu Arg
85 90 95

Glu Val Val Arg Gly Cys Ser Arg Phe Ala Gln Leu Ile Asn Ile Lys
100 105 110

His Glu Ile Asn Ala Asn Leu Leu Phe Ile Arg Phe Glu Phe Thr Thr
115 120 125

Gly Asp Ala Ser Gly His Asn Met Ala Thr Leu Ala Ser Asp Val Leu
130 135 140

Leu Gly His Leu Leu Glu Thr Ile Pro Gly Ile Ser Tyr Gly Ser Ile

145	150	155	160
Ser Gly Asn Tyr Cys Thr Asp Lys Lys Ala Thr Ala Ile Asn Gly Ile			
165	170	175	
Leu Gly Arg Gly Lys Asn Val Ile Thr Glu Leu Leu Val Pro Arg Asp			
180	185	190	
Val Val Glu Asn Asn Leu His Thr Thr Ala Ala Lys Ile Val Glu Leu			
195	200	205	
Asn Ile Arg Lys Asn Leu Leu Gly Thr Leu Leu Ala Gly Gly Ile Arg			
210	215	220	
Ser Ala Asn Ala His Phe Ala Asn Met Leu Leu Gly Phe Tyr Leu Ala			
225	230	235	240
Thr Gly Gln Asp Ala Ala Asn Ile Val Glu Gly Ser Gln Gly Val Val			
245	250	255	
Met Ala Glu Asp Arg Asp Gly Asp Leu Tyr Phe Ala Cys Thr Leu Pro			
260	265	270	
Asn Leu Ile Val Gly Thr Val Gly Asn Gly Lys Gly Leu Gly Phe Val			
275	280	285	
Glu Thr Asn Leu Ala Arg Leu Gly Cys Arg Ala Asp Arg Glu Pro Gly			
290	295	300	
Glu Asn Ala Arg Arg Leu Ala Val Ile Ala Ala Ala Thr Val Leu Cys			
305	310	315	320
Gly Glu Leu Ser Leu Leu Ala Ala Gln Thr Asn Pro Gly Glu Leu Met			
325	330	335	
Arg Ala His Val Gln Leu Glu Arg Asp Asn Lys Thr Ala Lys Val Gly			
340	345	350	

Ala

<210> 55

<211> 353

<212> PRT

<213> Streptomyces griseolosporeus

<400> 55

Met Thr Glu Ala His Ala Thr Ala Gly Val Pro Met Arg Trp Val Gly
1 5 10 15

Pro Val Arg Ile Ser Gly Asn Val Ala Thr Ile Glu Thr Gln Val Pro

20	25	30
Leu Ala Thr Tyr Glu Ser Pro Leu Trp Pro Ser Val Gly Arg Gly Ala		
35	40	45
Lys Val Ser Arg Leu Thr Glu Lys Gly Ile Val Ala Thr Leu Val Asp		
50	55	60
Glu Arg Met Thr Arg Ser Val Leu Val Glu Ala Thr Asp Ala Leu Thr		
65	70	75
80		
Ala Leu Ser Ala Ala Arg Thr Ile Glu Ala Arg Ile Asp Glu Leu Arg		
85	90	95
Glu Leu Val Arg Gly Cys Ser Arg Phe Ala Gln Leu Ile Gly Ile Arg		
100	105	110
His Glu Ile Thr Gly Asn Leu Leu Phe Val Arg Phe Glu Phe Ser Thr		
115	120	125
Gly Asp Ala Ser Gly His Asn Met Ala Thr Leu Ala Ser Asp Val Leu		
130	135	140
Leu Gln His Leu Leu Glu Thr Val Pro Gly Ile Ser Tyr Gly Ser Ile		
145	150	155
160		
Ser Gly Asn Tyr Cys Thr Asp Lys Lys Ala Thr Ala Ile Asn Gly Ile		
165	170	175
Leu Gly Arg Gly Lys Asn Val Val Thr Glu Leu Leu Val Pro Arg Asp		
180	185	190
Val Val Ala Asp Val Leu Asn Thr Thr Ala Ala Lys Ile Ala Glu Leu		
195	200	205
Asn Leu Arg Lys Asn Leu Leu Gly Thr Leu Leu Ala Gly Gly Ile Arg		
210	215	220
Ser Ala Asn Ala His Tyr Ala Asn Met Leu Leu Ala Phe Tyr Leu Ala		
225	230	235
240		
Thr Gly Gln Asp Ala Ala Asn Ile Val Glu Gly Ser Gln Gly Val Val		
245	250	255
Thr Ala Glu Asp Arg Asp Gly Asp Leu Tyr Leu Ala Cys Thr Leu Pro		
260	265	270
Asn Leu Ile Val Gly Thr Val Gly Asn Gly Lys Gly Leu Gly Phe Val		
275	280	285
Glu Thr Asn Leu Asn Arg Leu Gly Cys Arg Ala Asp Arg Glu Pro Gly		
290	295	300
Glu Asn Ala Arg Arg Leu Ala Val Ile Ala Ala Ala Thr Val Leu Cys		
305	310	315
320		
Gly Glu Leu Ser Leu Leu Ala Ala Gln Thr Asn Pro Gly Glu Leu Met		

325

330

335

Arg Ala His Val Gln Leu Glu Arg Gly His Thr Thr Ala Lys Ala Gly
 340 345 350

Val

<210> 56

<211> 353

<212> PRT

<213> Streptomyces sp. strain KO-3899

<400> 56

Met Thr Asp Thr His Ala Ile Ala Met Val Pro Met Lys Trp Val Gly
 1 5 10 15

Pro Leu Arg Ile Ser Gly Asn Val Ala Thr Thr Glu Thr His Val Pro
 20 25 30

Leu Ala Thr Tyr Glu Thr Pro Leu Trp Pro Ser Val Gly Arg Gly Ala
 35 40 45

Lys Val Ser Met Leu Ser Glu Arg Gly Ile Ala Ala Thr Leu Val Asp
 50 55 60

Glu Arg Met Thr Arg Ser Val Leu Val Glu Ala Thr Asp Ala Gln Thr
 65 70 75 80

Ala Tyr Thr Ala Ala Arg Ala Ile Glu Ala Arg Ile Glu Glu Leu Arg
 85 90 95

Ala Val Val Arg Thr Cys Ser Arg Phe Ala Glu Leu Leu Gln Val Arg
 100 105 110

His Glu Ile Ala Gly Asn Leu Leu Phe Val Arg Phe Glu Phe Ser Thr
 115 120 125

Arg Arg Pro Ser Gly His Asn Met Ala Thr Leu Ala Ser Asp Ala Leu
 130 135 140

Leu Ala His Leu Leu Gln Thr Ile Pro Gly Ile Ser Tyr Gly Ser Ile
 145 150 155 160

Ser Gly Asn Tyr Cys Thr Asp Lys Lys Ala Thr Ala Ile Asn Gly Ile
 165 170 175

Leu Gly Arg Gly Lys Asn Val Val Thr Glu Leu Val Val Pro Arg Glu
 180 185 190

Val Val Glu Arg Val Leu His Thr Thr Ala Ala Lys Ile Val Glu Leu

195	200	205
Asn Ile Arg Lys Asn Leu Leu Gly Thr Leu Leu Ala Gly Gly Ile Arg		
210	215	220
Ser Ala Asn Ala His Tyr Ala Asn Met Leu Leu Gly Phe Tyr Leu Ala		
225	230	235
240		
Thr Gly Gln Asp Ala Ala Asn Ile Val Glu Gly Ser Gln Gly Val Thr		
245	250	255
Leu Ala Glu Asp Arg Asp Gly Asp Leu Tyr Phe Ser Cys Asn Leu Pro		
260	265	270
Asn Leu Ile Val Gly Thr Val Gly Asn Gly Lys Gly Leu Glu Phe Val		
275	280	285
Glu Thr Asn Leu Asn Arg Leu Gly Cys Arg Glu Asp Arg Ala Pro Gly		
290	295	300
Glu Asn Ala Arg Arg Leu Ala Val Ile Ala Ala Ala Thr Val Leu Cys		
305	310	315
320		
Gly Glu Leu Ser Leu Leu Ala Ala Gln Thr Asn Pro Gly Glu Leu Met		
325	330	335
Arg Ala His Val Glu Leu Glu Arg Asp Asn Thr Thr Ala Glu Val Gly		
340	345	350
Val		
<210> 57		
<211> 347		
<212> PRT		
<213> Erwinia herbicola		
<400> 57		
Met Lys Asp Glu Arg Leu Val Gln Arg Lys Asn Asp His Leu Asp Ile		
1	5	10
15		
Val Leu Asp Pro Arg Arg Ala Val Thr Gln Ala Ser Ala Gly Phe Glu		
20	25	30
Arg Trp Arg Phe Thr His Cys Ala Leu Pro Glu Leu Asn Phe Ser Asp		
35	40	45
Ile Thr Leu Glu Thr Thr Phe Leu Asn Arg Gln Leu Gln Ala Pro Leu		
50	55	60
Leu Ile Ser Ser Met Thr Gly Gly Val Glu Arg Ser Arg His Ile Asn		

65	70	75	80
Arg His Leu Ala Glu Ala Ala Gln Val		Leu Lys Ile Ala Met Gly Val	
85		90	95
Gly Ser Gln Arg Val Ala Ile Glu Ser Asp Ala Gly	Leu Gly	Leu Asp	
100	105	110	
Lys Thr Leu Arg Gln Leu Ala Pro Asp Val	Pro Leu	Leu Ala Asn Leu	
115	120	125	
Gly Ala Ala Gln Leu Thr Gly Arg Lys Gly Ile Asp	Tyr Ala Arg Arg		
130	135	140	
Ala Val Glu Met Ile Glu Ala Asp Ala Leu Ile Val	His Leu Asn Pro		
145	150	155	160
Leu Gln Glu Ala Leu Gln Pro Gly Gly Asp Arg Asp	Trp Arg Gly Arg		
165	170	175	
Leu Ala Ala Ile Glu Thr Leu Val Arg Glu Leu Pro	Val Pro Leu Val		
180	185	190	
Val Lys Glu Val Gly Ala Gly Ile Ser Arg Thr Val	Ala Gly Gln Leu		
195	200	205	
Ile Asp Ala Gly Val Thr Val Ile Asp Val Ala Gly	Ala Gly Gly Thr		
210	215	220	
Ser Trp Ala Ala Val Glu Gly Glu Arg Ala Ala	Thr Glu Gln Gln Arg		
225	230	235	240
Ser Val Ala Asn Val Phe Ala Asp Trp Gly Ile Pro	Thr Ala Glu Ala		
245	250	255	
Leu Val Asp Ile Ala Glu Ala Trp Pro Gln Met Pro	Leu Ile Ala Ser		
260	265	270	
Gly Gly Ile Lys Asn Gly Val Asp Ala Ala Lys Ala	Leu Arg Leu Gly		
275	280	285	
Ala Cys Met Val Gly Gln Ala Ala Ala Val Leu	Gly Ser Ala Gly Val		
290	295	300	
Ser Thr Glu Lys Val Ile Asp His Phe Asn Val	Ile Ile Glu Gln Leu		
305	310	315	320
Arg Val Ala Cys Phe Cys Thr Gly Ser Arg Ser	Leu Ser Asp Leu Lys		
325	330	335	
Gln Ala Asp Ile Arg Tyr Val Arg Asp Thr Pro			
340	345		
<210> 58			
<211> 360			

<212> PRT

<213> *Borrelia burgdorferi*

<400> 58

Met Met Asp Thr Glu Phe Met Gly Ile Glu Pro Asn Ile Leu Glu Asn
1 5 10 15

Lys Lys Arg His Ile Glu Ile Cys Leu Asn Lys Asn Asp Val Lys Gly
20 25 30

Gly Cys Asn Phe Leu Lys Phe Ile Lys Leu Lys His Asn Ala Leu Ser
35 40 45

Asp Phe Asn Phe Ser Glu Ile Asn Ile Lys Glu Glu Ile Phe Gly Tyr
50 55 60

Asn Ile Ser Met Pro Val Phe Ile Ser Ser Met Thr Gly Gly Ser Lys
65 70 75 80

Glu Gly Asn Asp Phe Asn Lys Ser Leu Val Arg Ile Ala Asn Tyr Leu
85 90 95

Lys Ile Pro Ile Gly Leu Gly Ser Phe Lys Leu Leu Phe Lys Tyr Pro
100 105 110

Glu Tyr Ile Arg Asp Phe Thr Leu Lys Arg Tyr Ala His Asn Ile Pro
115 120 125

Leu Phe Ala Asn Val Gly Ala Val Gln Ile Val Glu Phe Gly Ile Ser
130 135 140

Lys Ile Ala Glu Met Ile Lys Arg Leu Glu Val Asp Ala Ile Ile Val
145 150 155 160

His Leu Asn Ala Gly Gln Glu Leu Met Lys Val Asp Gly Asp Arg Asn
165 170 175

Phe Lys Gly Ile Arg Glu Ser Ile Ala Lys Leu Ser Asp Phe Leu Ser
180 185 190

Val Pro Leu Ile Val Lys Glu Thr Gly Phe Gly Ile Ser Pro Lys Asp
195 200 205

Val Lys Glu Leu Phe Ser Leu Gly Ala Ser Tyr Val Asp Leu Ala Gly
210 215 220

Ser Gly Gly Thr Asn Trp Ile Leu Val Glu Gly Met Lys Ser Asn Asn
225 230 235 240

Leu Asn Ile Ala Ser Cys Phe Ser Asp Trp Gly Ile Pro Ser Val Phe
245 250 255

Thr Leu Leu Ser Ile Asp Asp Ser Leu Lys Ala Asn Ile Phe Ala Ser

260

265

270

Gly Gly Tyr Glu Thr Gly Met Asp Ile Ala Lys Gly Ile Ala Leu Gly
 275 280 285

Ala Arg Leu Ile Gly Val Ala Ala Val Val Leu Arg Ala Phe Tyr Asp
 290 295 300

Ser Gly Glu Asp Ala Val Phe Gly Leu Phe Ser Asp Tyr Glu His Ile
 305 310 315 320

Leu Lys Met Ser Met Phe Leu Ser Gly Ser Lys Ser Leu Leu Glu Phe
 325 330 335

Arg Asn Asn Lys Tyr Phe Leu Ser Ser Tyr Leu Leu Asp Glu Leu Gly
 340 345 350

Val Phe Lys Gln Phe Tyr Gly Thr
 355 360

<210> 59

<211> 349

<212> PRT

<213> Synechocystis sp. PCC 6803

<400> 59

Met Asp Ser Thr Pro His Arg Lys Ser Asp His Ile Arg Ile Val Leu
 1 5 10 15

Glu Glu Asp Val Val Gly Lys Gly Ile Ser Thr Gly Phe Glu Arg Leu
 20 25 30

Met Leu Glu His Cys Ala Leu Pro Ala Val Asp Leu Asp Ala Val Asp
 35 40 45

Leu Gly Leu Thr Leu Trp Gly Lys Ser Leu Thr Tyr Pro Trp Leu Ile
 50 55 60

Ser Ser Met Thr Gly Gly Thr Pro Glu Ala Lys Gln Ile Asn Leu Phe
 65 70 75 80

Leu Ala Glu Val Ala Gln Ala Leu Gly Ile Ala Met Gly Leu Gly Ser
 85 90 95

Gln Arg Ala Ala Ile Glu Asn Pro Asp Leu Ala Phe Thr Tyr Gln Val
 100 105 110

Arg Ser Val Ala Pro Asp Ile Leu Leu Phe Ala Asn Leu Gly Leu Val
 115 120 125

Gln Leu Asn Tyr Gly Tyr Gly Leu Glu Gln Ala Gln Arg Ala Val Asp

130	135	140
Met Ile Glu Ala Asp Ala Leu Ile Leu His Leu Asn Pro Leu Gln Glu		
145	150	155
Ala Val Gln Pro Asp Gly Asp Arg Leu Trp Ser Gly Leu Trp Ser Lys		
165	170	175
Leu Glu Ala Leu Val Glu Ala Leu Glu Val Pro Val Ile Val Lys Glu		
180	185	190
Val Gly Asn Gly Ile Ser Gly Pro Val Ala Lys Arg Leu Gln Glu Cys		
195	200	205
Gly Val Gly Ala Ile Asp Val Ala Gly Ala Gly Thr Ser Trp Ser		
210	215	220
Glu Val Glu Ala His Arg Gln Thr Asp Arg Gln Ala Lys Glu Val Ala		
225	230	235
His Asn Phe Ala Asp Trp Gly Leu Pro Thr Ala Trp Ser Leu Gln Gln		
245	250	255
Val Val Gln Asn Thr Glu Gln Ile Leu Val Phe Ala Ser Gly Gly Ile		
260	265	270
Arg Ser Gly Ile Asp Gly Ala Lys Ala Ile Ala Leu Gly Ala Thr Leu		
275	280	285
Val Gly Ser Ala Ala Pro Val Leu Ala Glu Ala Lys Ile Asn Ala Gln		
290	295	300
Arg Val Tyr Asp His Tyr Gln Ala Arg Leu Arg Glu Leu Gln Ile Ala		
305	310	315
Ala Phe Cys Cys Asp Ala Ala Asn Leu Thr Gln Leu Ala Gln Val Pro		
325	330	335
Leu Trp Asp Arg Gln Ser Gly Gln Arg Leu Thr Lys Pro		
340	345	
<210> 60		
<211> 361		
<212> PRT		
<213> Streptomyces sp. CL190		
<400> 60		
Met Thr Ser Ala Gln Arg Lys Asp Asp His Val Arg Leu Ala Ile Glu		
1	5	10
Gln His Asn Ala His Ser Gly Arg Asn Gln Asp Asp Val Ser Phe Val		

20

25

30

His His Ala Leu Ala Gly Ile Asp Arg Pro Asp Val Ser Leu Ala Thr
 35 40 45

Ser Phe Ala Gly Ile Ser Trp Gln Val Pro Ile Tyr Ile Asn Ala Met
 50 55 60

Thr Gly Gly Ser Glu Lys Thr Gly Leu Ile Asn Arg Asp Leu Ala Thr
 65 70 75 80

Ala Ala Arg Glu Thr Gly Val Pro Ile Ala Ser Gly Ser Met Asn Ala
 85 90 95

Tyr Ile Lys Asp Pro Cys Ala Asp Thr Phe Arg Val Leu Arg Asp Glu
 100 105 110

Asn Pro Asn Gly Phe Val Ile Ala Asn Ile Asn Ala Thr Thr Thr Val
 115 120 125

Asp Asn Ala Gln Arg Ala Ile Asp Leu Ile Glu Ala Asn Ala Leu Gln
 130 135 140

Ile His Ile Asn Thr Ala Gln Glu Thr Pro Met Pro Glu Gly Asp Arg
 145 150 155 160

Ser Phe Ala Ser Trp Val Pro Gln Ile Glu Lys Ile Ala Ala Ala Val
 165 170 175

Asp Ile Pro Val Ile Val Lys Glu Val Gly Asn Gly Leu Ser Arg Gln
 180 185 190

Thr Ile Leu Leu Leu Ala Asp Leu Gly Val Gln Ala Ala Asp Val Ser
 195 200 205

Gly Arg Gly Gly Thr Asp Phe Ala Arg Ile Glu Asn Gly Arg Arg Glu
 210 215 220

Leu Gly Asp Tyr Ala Phe Leu His Gly Trp Gly Gln Ser Thr Ala Ala
 225 230 235 240

Cys Leu Leu Asp Ala Gln Asp Ile Ser Leu Pro Val Leu Ala Ser Gly
 245 250 255

Gly Val Arg His Pro Leu Asp Val Val Arg Ala Leu Ala Leu Gly Ala
 260 265 270

Arg Ala Val Gly Ser Ser Ala Gly Phe Leu Arg Thr Leu Met Asp Asp
 275 280 285

Gly Val Asp Ala Leu Ile Thr Lys Leu Thr Thr Trp Leu Asp Gln Leu
 290 295 300

Ala Ala Leu Gln Thr Met Leu Gly Ala Arg Thr Pro Ala Asp Leu Thr
 305 310 315 320

Arg Cys Asp Val Leu Leu His Gly Glu Leu Arg Asp Phe Cys Ala Asp

325 330 335

Arg Gly Ile Asp Thr Arg Arg Leu Ala Gln Arg Ser Ser Ser Ile Glu
340 345 350

Ala Leu Gln Thr Thr Gly Ser Thr Arg
355 360

<210> 61

<211> 364

<212> PRT

<213> *Streptomyces griseolosporeus*

<400> 61

Met Ser Ser Ala Gln Arg Lys Asp Asp His Val Arg Leu Ala Thr Glu
1 5 10 15

Gln Gln Arg Ala His Ser Gly Arg Asn Gln Phe Asp Asp Val Ser Phe
20 25 30

Val His His Ala Leu Ala Gly Ile Asp Arg Pro Asp Val Arg Leu Ala
35 40 45

Thr Thr Phe Ala Gly Ile Thr Trp Arg Leu Pro Leu Tyr Ile Asn Ala
50 55 60

Met Thr Gly Gly Ser Ala Lys Thr Gly Ala Ile Asn Arg Asp Leu Ala
65 70 75 80

Val Ala Ala Arg Glu Thr Gly Ala Ala Ile Ala Ser Gly Ser Met His
85 90 95

Ala Phe Phe Arg Asp Pro Ser Cys Ala Asp Thr Phe Arg Val Leu Arg
100 105 110

Thr Glu Asn Pro Asp Gly Phe Val Met Ala Asn Val Asn Ala Thr Ala
115 120 125

Ser Val Asp Asn Ala Arg Arg Ala Val Asp Leu Ile Glu Ala Asn Ala
130 135 140

Leu Gln Ile His Leu Asn Thr Ala Gln Glu Thr Pro Met Pro Glu Gly
145 150 155 160

Asp Arg Ser Phe Gly Ser Trp Pro Ala Gln Ile Ala Lys Ile Thr Ala
165 170 175

Ala Val Asp Val Pro Val Ile Val Lys Glu Val Gly Asn Gly Leu Ser
180 185 190

Arg Gln Thr Leu Leu Ala Leu Pro Asp Leu Gly Val Arg Val Ala Asp

195	200	205
Val Ser Gly Arg Gly Gly	Thr Asp Phe Ala Arg	Ile Glu Asn Ser Arg
210	215	220
Arg Pro Leu Gly Asp Tyr Ala Phe Leu His	Gly Trp Gly Gln Ser Thr	240
225	230	235
Pro Ala Cys Leu Leu Asp Ala Gln Asp Val	Gly Phe Pro Leu Leu Ala	255
245	250	
Ser Gly Gly Ile Arg Asn Pro Leu Asp Val	Ala Arg Ala Leu Ala Leu	270
260	265	
Gly Ala Gly Ala Val Gly Ser Ser Gly Val	Phe Leu Arg Thr Leu Ile	285
275	280	
Asp Gly Gly Val Ser Ala Leu Val Ala Gln Ile	Ser Thr Trp Leu Asp	
290	295	300
Gln Leu Ala Ala Leu Gln Thr Met Leu Gly Ala Arg	Thr Pro Ala Asp	320
305	310	315
Leu Thr Arg Cys Asp Val Leu Ile His	Gly Pro Leu Arg Ser Phe Cys	335
325	330	
Thr Asp Arg Gly Ile Asp Ile Gly Arg Phe Ala Arg	Arg Ser Ser Ser	350
340	345	
Ala Asp Ile Arg Ser Glu Met Thr Gly Ser Thr Arg		
355	360	
<210> 62		
<211> 368		
<212> PRT		
<213> <i>Sulfolobus solfataricus</i>		

<400> 62		
Met Pro Asp Ile Val Asn Arg Lys Val Glu His Val Glu Ile Ala Ala		
1	5	10
15		
Phe Glu Asn Val Asp Gly Leu Ser Ser Ser Thr Phe Leu Asn Asp Val		
20	25	30
30		
Ile Leu Val His Gln Gly Phe Pro Gly Ile Ser Phe Ser Glu Ile Asn		
35	40	45
45		
Thr Lys Thr Lys Phe Phe Arg Lys Glu Ile Ser Ala Pro Ile Met Val		
50	55	60
60		
Thr Gly Met Thr Gly Gly Arg Asn Glu Leu Gly Arg Ile Asn Arg Ile		

65	70	75	80
Ile Ala Glu Val Ala Glu Lys Phe Gly Ile Pro Met Gly Val Gly Ser			
85	90	95	
Gln Arg Val Ala Ile Glu Lys Ala Glu Ala Arg Glu Ser Phe Thr Ile			
100	105	110	
Val Arg Lys Val Ala Pro Thr Ile Pro Ile Ile Ala Asn Leu Gly Met			
115	120	125	
Pro Gln Leu Val Lys Gly Tyr Gly Leu Lys Glu Phe Gln Asp Ala Ile			
130	135	140	
Gln Met Ile Glu Ala Asp Ala Ile Ala Val His Leu Asn Pro Ala Gln			
145	150	155	160
Glu Val Phe Gln Pro Glu Gly Glu Pro Glu Tyr Gln Ile Tyr Ala Leu			
165	170	175	
Glu Arg Leu Arg Asp Ile Ser Lys Glu Leu Ser Val Pro Ile Ile Val			
180	185	190	
Lys Glu Ser Gly Asn Gly Ile Ser Met Glu Thr Ala Lys Leu Leu Tyr			
195	200	205	
Ser Tyr Gly Ile Lys Asn Phe Asp Thr Ser Gly Gln Gly Gly Thr Asn			
210	215	220	
Trp Ile Ala Ile Glu Met Ile Arg Asp Ile Arg Arg Gly Asn Trp Lys			
225	230	235	240
Ala Glu Ser Ala Lys Asn Phe Leu Asp Trp Gly Val Pro Thr Ala Ala			
245	250	255	
Ser Ile Ile Glu Val Arg Tyr Ser Ile Pro Asp Ala Phe Leu Val Gly			
260	265	270	
Ser Gly Gly Ile Arg Ser Gly Leu Asp Ala Ala Lys Ala Ile Ala Leu			
275	280	285	
Gly Ala Asp Ile Ala Gly Met Ala Leu Pro Val Leu Lys Ser Ala Ile			
290	295	300	
Glu Gly Lys Glu Ser Leu Glu Gln Phe Phe Arg Lys Ile Ile Phe Glu			
305	310	315	320
Leu Lys Ala Thr Met Met Leu Thr Gly Ser Lys Asn Val Glu Ala Leu			
325	330	335	
Lys Arg Ser Ser Ile Val Ile Leu Gly Lys Leu Lys Glu Trp Ala Glu			
340	345	350	
Tyr Arg Gly Ile Asn Leu Ser Ile Tyr Glu Lys Val Arg Lys Arg Glu			
355	360	365	

<210> 63

<211> 342

<212> PRT

<213> Rickettsia prowazekii

<400> 63

Met Pro Lys Glu Gln Asn Leu Asp Ile Glu Arg Lys Gln Glu His Ile
1 5 10 15

Glu Ile Asn Leu Lys Gln Asn Val Asn Ser Thr Leu Lys Ser Gly Leu
20 25 30

Glu Ser Ile Lys Phe Ile His Asn Ala Leu Pro Glu Ile Asn Tyr Asp
35 40 45

Ser Ile Asp Thr Thr Thr Phe Leu Gly Lys Asp Met Lys Ala Pro
50 55 60

Ile Leu Ile Ser Ser Met Thr Gly Gly Thr Ala Arg Ala Arg Asp Ile
65 70 75 80

Asn Tyr Arg Leu Ala Gln Ala Ala Gln Lys Ser Gly Ile Ala Met Gly
85 90 95

Leu Gly Ser Met Arg Ile Leu Leu Thr Lys Pro Asp Thr Ile Lys Thr
100 105 110

Phe Thr Val Arg His Val Ala Pro Asp Ile Pro Leu Leu Ala Asn Ile
115 120 125

Gly Ala Val Gln Leu Asn Tyr Gly Val Thr Pro Lys Glu Cys Gln Tyr
130 135 140

Leu Ile Asp Thr Ile Lys Ala Asp Ala Leu Ile Leu His Leu Asn Val
145 150 155 160

Leu His Glu Leu Thr Gln Pro Glu Gly Asn Lys Asn Trp Glu Asn Leu
165 170 175

Leu Pro Lys Ile Lys Glu Val Ile Asn Tyr Leu Ser Val Pro Val Ile
180 185 190

Val Lys Glu Val Gly Tyr Gly Leu Ser Lys Gln Val Ala Lys Lys Leu
195 200 205

Ile Lys Ala Gly Val Lys Val Leu Asp Ile Ala Gly Ser Gly Gly Thr
210 215 220

Ser Trp Ser Gln Val Glu Ala Tyr Arg Ala Lys Asn Ser Met Gln Asn
225 230 235 240

Arg Ile Ala Ser Ser Phe Ile Asn Trp Gly Ile Thr Thr Leu Asp Ser

245 250 255

Leu Lys Met Leu Gln Glu Ile Ser Lys Asp Ile Thr Ile Ile Ala Ser
260 265 270

Gly Gly Leu Gln Ser Gly Ile Asp Gly Ala Lys Ala Ile Arg Met Gly
275 280 285

Ala Asn Ile Phe Gly Leu Ala Gly Lys Leu Leu Lys Ala Ala Asp Ile
290 295 300

Ala Glu Ser Leu Val Leu Glu Glu Ile Gln Val Ile Ile Glu Gln Leu
305 310 320

Lys Ile Thr Met Leu Cys Thr Gly Ser Cys Thr Leu Lys Asp Leu Ala
325 330 335

Lys Ala Glu Ile Met Trp
340

<210> 64

<211> 286

<212> PRT

<213> Deinococcus radiodurans

<400> 64

Met Arg Leu Asp Thr Val Phe Leu Gly Arg Arg Leu Lys Ala Pro Val
1 5 10 15

Leu Ile Gly Ala Met Thr Gly Gly Ala Glu Lys Ala Gly Val Ile Asn
20 25 30

Arg Asn Leu Ala Thr Ala Ala Arg Asn Leu Gly Leu Gly Met Met Leu
35 40 45

Gly Ser Gln Arg Val Met Leu Glu His Pro Asp Ala Trp Glu Ser Phe
50 55 60

Asn Val Arg Glu Val Ala Pro Glu Ile Leu Leu Ile Gly Asn Leu Gly
65 70 75 80

Ala Ala Gln Phe Met Leu Gly Tyr Gly Ala Glu Gln Ala Arg Arg Ala
85 90 95

Val Asp Glu Val Met Ala Asp Ala Leu Ala Ile His Leu Asn Pro Leu
100 105 110

Gln Glu Ala Leu Gln Arg Gly Gly Asp Thr Arg Trp Gln Gly Val Thr
115 120 125

Tyr Arg Leu Lys Gln Val Ala Arg Glu Leu Asp Phe Pro Val Ile Ile

130	135	140
Lys Glu Val Gly His Gly Leu Asp Ala Ala Thr Leu Arg Ala Leu Ala		
145	150	155
Asp Gly Pro Phe Ala Ala Tyr Asp Val Ala Gly Ala Gly Gly Thr Ser		
165	170	175
Trp Ala Arg Val Glu Gln Leu Val Ala His Gly Gln Val His Ser Pro		
180	185	190
Asp Leu Cys Glu Leu Gly Val Pro Thr Ala Gln Ala Leu Arg Gln Ala		
195	200	205
Arg Lys Thr Leu Pro Gly Ala Gln Leu Ile Ala Ser Gly Gly Ile Arg		
210	215	220
Ser Gly Leu Asp Ala Ala Arg Ala Leu Ser Leu Gly Ala Glu Val Val		
225	230	235
Ala Val Ala Arg Pro Leu Leu Glu Pro Ala Leu Asp Ser Ser Glu Ala		
245	250	255
Ala Glu Ala Trp Leu Arg Asn Phe Ile Gln Glu Leu Arg Val Ala Leu		
260	265	270
Phe Val Gly Gly Tyr Arg Asp Val Arg Glu Val Arg Gly Gly		
275	280	285
<210> 65		
<211> 361		
<212> PRT		
<213> Aeropyrum pernix		
C		
<400> 65		
Met Ile Val Ser Ser Lys Val Glu Ser Arg Glu Ser Thr Leu Leu Glu		
1	5	10
Tyr Val Arg Ile Val His Asn Pro Thr Pro Glu Val Asn Leu Gly Asp		
20	25	30
Val Ser Leu Glu Ile Asp Phe Cys Gly Gly Arg Leu Arg Ala Pro Leu		
35	40	45
Val Ile Thr Gly Met Thr Gly Gly His Pro Asp Val Glu Trp Ile Asn		
50	55	60
Arg Glu Leu Ala Ser Val Ala Glu Glu Leu Gly Ile Ala Ile Gly Val		
65	70	75
Gly Ser Gln Arg Ala Ala Ile Glu Asp Pro Ser Leu Ala Arg Thr Phe		

85	90	95
Arg Ala Ala Arg Glu Ala Ala Pro Asn Ala Phe Leu Ile Ala Asn Leu		
100	105	110
Gly Ala Pro Gln Leu Ser Leu Gly Tyr Ser Val Arg Glu Val Arg Met		
115	120	125
Ala Val Glu Met Ile Asp Ala Asp Ala Ile Ala Ile His Leu Asn Pro		
130	135	140
Gly Gln Glu Ala Tyr Gln Pro Glu Gly Asp Pro Phe Tyr Arg Gly Val		
145	150	155
160		
Val Gly Lys Ile Ala Glu Ala Ala Glu Ala Ala Gly Val Pro Val Ile		
165	170	175
Val Lys Glu Thr Gly Asn Gly Leu Ser Arg Glu Ala Val Ala Gln Leu		
180	185	190
Arg Ala Leu Gly Val Arg Cys Phe Asp Val Ala Gly Leu Gly Thr		
195	200	205
Asn Trp Ile Lys Ile Glu Val Leu Arg Gly Arg Lys Ala Gly Ser Pro		
210	215	220
Leu Glu Ala Gly Pro Leu Gln Asp Phe Trp Gly Asn Pro Thr Ala Ala		
225	230	235
240		
Ala Leu Met Glu Ala Arg Thr Ala Ala Pro Asp Ala Tyr Ile Ile Ala		
245	250	255
Ser Gly Gly Val Arg Asn Gly Leu Asp Ala Ala Arg Ala Ile Ala Leu		
260	265	270
Gly Ala Asp Ala Ala Gly Val Ala Leu Pro Ala Ile Arg Ser Leu Leu		
275	280	285
Ser Gly Gly Arg Gln Ala Thr Leu Lys Leu Leu Lys Ala Ile Glu Tyr		
290	295	300
Gln Leu Lys Thr Ala Val Tyr Met Val Gly Glu Thr Arg Val Arg Gly		
305	310	315
320		
Leu Trp Arg Ala Pro Ile Val Val Trp Gly Arg Leu Ala Glu Glu Ala		
325	330	335
Glu Ala Arg Gly Ile Asp Pro Arg Trp Tyr Thr Asn Thr Leu Arg Leu		
340	345	350
Glu Ala Leu Val Tyr Lys Asp Val Lys		
355	360	
<210> 66		
<211> 379		

<212> PRT

<213> Halobacterium sp. NRC-1

<400> 66

Met Gly Glu Ser Arg Tyr Asn Ser Ile Val Phe Pro Ser Leu Val Gln
1 5 10 15

Thr Arg Leu Met Thr Ala Gln Asp Ser Thr Gln Thr Glu Asp Arg Lys
20 25 30

Asp Asp His Leu Gln Ile Val Gln Glu Arg Asp Val Glu Thr Thr Gly
35 40 45

Thr Gly Phe Asp Asp Val His Leu Val His Asn Ala Leu Pro Glu Leu
50 55 60

Asp Tyr Asp Ala Ile Asp Pro Ser Ile Asp Phe Leu Gly His Asp Leu
65 70 75 80

Ser Ala Pro Ile Phe Ile Glu Ser Met Thr Gly Gly His His Asn Thr
85 90 95

Thr Glu Ile Asn Arg Ala Leu Ala Arg Ala Ala Ser Glu Thr Gly Ile
100 105 110

Ala Met Gly Leu Gly Ser Gln Arg Ala Gly Leu Glu Leu Asp Asp Glu
115 120 125

Arg Val Leu Glu Ser Tyr Thr Val Val Arg Asp Ala Ala Pro Asp Ala
130 135 140

Phe Ile Tyr Gly Asn Leu Gly Ala Ala Gln Leu Arg Glu Tyr Asp Ile
145 150 155 160

Glu Met Val Glu Gln Ala Val Glu Met Ile Asp Ala Asp Ala Leu Ala
165 170 175

Val His Leu Asn Phe Leu Gln Glu Ala Thr Gln Pro Glu Gly Asp Val
180 185 190

Asp Gly Arg Asn Cys Val Ala Ala Ile Glu Arg Val Ser Glu Ala Leu
195 200 205

Ser Val Pro Ile Ile Val Lys Glu Thr Gly Asn Gly Ile Ser Gly Glu
210 215 220

Thr Ala Arg Glu Leu Thr Ala Ala Gly Val Asp Ala Leu Asp Val Ala
225 230 235 240

Gly Lys Gly Gly Thr Thr Trp Ser Gly Ile Glu Ala Tyr Arg Ala Ala
245 250 255

Ala Ala Asn Ala Pro Arg Gln Lys Gln Ile Gly Thr Leu Phe Arg Glu

260

265

270

Trp Gly Ile Pro Thr Ala Ala Ser Thr Ile Glu Cys Val Ala Glu His
 275 280 285

Asp Cys Val Ile Ala Ser Gly Gly Val Arg Thr Gly Leu Asp Val Ala
 290 295 300

Lys Ala Ile Ala Leu Gly Ala Arg Ala Gly Gly Leu Ala Lys Pro Phe
 305 310 315 320

Leu Lys Pro Ala Thr Asp Gly Pro Asp Ala Val Ile Glu Arg Val Gly
 325 330 335

Asp Leu Ile Ala Glu Leu Arg Thr Ala Met Phe Val Thr Gly Ser Gly
 340 345 350

Ser Ile Asp Glu Leu Gln Gln Val Glu Tyr Val Leu His Gly Lys Thr
 355 360 365

Arg Glu Tyr Val Glu Gln Arg Thr Ser Ser Glu
 370 375

<210> 67

<211> 317

<212> PRT

<213> Archaeoglobus fulgidus

<400> 67

Met Met Leu Ile His Lys Ala Leu Pro Glu Val Asp Tyr Trp Lys Ile
 1 5 10 15

Asp Thr Glu Ile Glu Phe Phe Gly Lys Lys Leu Ser Phe Pro Leu Leu
 20 25 30

Ile Ala Ser Met Thr Gly Gly His Pro Glu Thr Lys Glu Ile Asn Ala
 35 40 45

Arg Leu Gly Glu Ala Val Glu Glu Ala Gly Ile Gly Met Gly Val Gly
 50 55 60

Ser Gln Arg Ala Ala Ile Glu Asp Glu Ser Leu Ala Asp Ser Phe Thr
 65 70 75 80

Val Val Arg Glu Lys Ala Pro Asn Ala Phe Val Tyr Ala Asn Ile Gly
 85 90 95

Met Pro Gln Val Ile Glu Arg Gly Val Glu Ile Val Asp Arg Ala Val
 100 105 110

Glu Met Ile Asp Ala Asp Ala Val Ala Ile His Leu Asn Tyr Leu Gln

115	120	125
Glu Ala Ile Gln Pro Glu Gly Asp Leu Asn Ala Glu Lys Gly Leu Glu		
130	135	140
Val Leu Glu Glu Val Cys Arg Ser Val Lys Val Pro Val Ile Ala Lys		
145	150	155
Glu Thr Gly Ala Gly Ile Ser Arg Glu Val Ala Val Met Leu Lys Arg		
165	170	175
Ala Gly Val Ser Ala Ile Asp Val Gly Gly Lys Gly Thr Thr Phe		
180	185	190
Ser Gly Val Glu Val Tyr Arg Val Asn Asp Glu Val Ser Lys Ser Val		
195	200	205
Gly Ile Asp Phe Trp Asp Trp Gly Leu Pro Thr Ala Phe Ser Ile Val		
210	215	220
Asp Cys Arg Gly Ile Leu Pro Val Ile Ala Thr Gly Gly Leu Arg Ser		
225	230	235
Gly Leu Asp Val Ala Lys Ser Ile Ala Ile Gly Ala Glu Leu Gly Ser		
245	250	255
Ala Ala Leu Pro Phe Leu Arg Ala Ala Val Glu Ser Ala Glu Lys Val		
260	265	270
Arg Glu Glu Ile Glu Tyr Phe Arg Arg Gly Leu Lys Thr Ala Met Phe		
275	280	285
Leu Thr Gly Cys Lys Asn Val Glu Glu Leu Lys Gly Leu Lys Val Phe		
290	295	300
Val Ser Gly Arg Leu Lys Glu Trp Ile Asp Phe Arg Gly		
305	310	315
<210> 68		
<211> 370		
<212> PRT		
<213> Pyrococcus abyssi		
<400> 68		
Met Glu Glu Gln Thr Ile Leu Arg Lys Phe Glu His Ile Lys His Cys		
1	5	10
Leu Thr Lys Asn Val Glu Ala His Val Thr Asn Gly Phe Glu Asp Val		
20	25	30
His Leu Ile His Lys Ser Leu Pro Glu Ile Asp Lys Asp Glu Ile Asp		

35	40	45
Leu Ser Val Lys Phe Leu Gly Arg Lys Phe Asp Tyr Pro Ile Met Ile		
50	55	60
Thr Gly Met Thr Gly Gly Thr Arg Lys Gly Glu Ile Ala Trp Arg Ile		
65	70	75
Asn Arg Thr Leu Ala Gln Ala Ala Gln Glu Leu Asn Ile Pro Leu Gly		
85	90	95
Leu Gly Ser Gln Arg Ala Met Ile Glu Lys Pro Glu Thr Trp Glu Ser		
100	105	110
Tyr Tyr Val Arg Asp Val Ala Pro Asp Val Phe Leu Val Gly Asn Leu		
115	120	125
Gly Ala Pro Gln Phe Gly Arg Asn Ala Lys Lys Arg Tyr Ser Val Asp		
130	135	140
Glu Val Leu Tyr Ala Ile Glu Lys Ile Glu Ala Asp Ala Ile Ala Ile		
145	150	155
His Met Asn Pro Leu Gln Glu Ser Ile Gln Pro Glu Gly Asp Thr Thr		
165	170	175
Phe Ser Gly Val Leu Glu Ala Leu Ala Glu Ile Thr Ser Thr Ile Asp		
180	185	190
Tyr Pro Val Ile Ala Lys Glu Thr Gly Ala Gly Val Ser Lys Glu Val		
195	200	205
Ala Val Glu Leu Glu Ala Val Gly Val Asp Ala Ile Asp Ile Ser Gly		
210	215	220
Leu Gly Gly Thr Ser Trp Ser Ala Val Glu Tyr Tyr Arg Thr Lys Asp		
225	230	235
Gly Glu Lys Arg Asn Leu Ala Leu Lys Phe Trp Asp Trp Gly Ile Lys		
245	250	255
Thr Ala Ile Ser Leu Ala Glu Val Arg Trp Ala Thr Asn Leu Pro Ile		
260	265	270
Ile Ala Ser Gly Gly Met Arg Asp Gly Ile Thr Met Ala Lys Ala Leu		
275	280	285
Ala Met Gly Ala Ser Met Val Gly Ile Ala Leu Pro Val Leu Arg Pro		
290	295	300
Ala Ala Lys Gly Asp Val Glu Gly Val Ile Arg Ile Ile Lys Gly Tyr		
305	310	315
Ala Glu Glu Ile Arg Asn Val Met Phe Leu Val Gly Ala Arg Asn Ile		
325	330	335
Lys Glu Leu Arg Lys Val Pro Leu Val Ile Thr Gly Phe Val Arg Glu		

340

345

350

Trp Leu Leu Gln Arg Ile Asp Leu Asn Ser Tyr Leu Arg Ala Arg Phe
355 360 365

Lys Met
370

<210> 69

<211> 371

<212> PRT

<400> 69

Met Lys Glu Glu Leu Thr Ile Leu Arg Lys Phe Glu His Ile Glu His
1 5 10 15

Cys Leu Lys Arg Asn Val Glu Ala His Val Ser Asn Gly Phe Glu Asp
20 25 30

Val Tyr Phe Val His Lys Ser Leu Pro Glu Ile Asp Lys Asp Glu Ile
35 40 45

Asp Leu Thr Val Glu Phe Leu Gly Arg Lys Phe Asp Tyr Pro Ile Met
50 55 60

Ile Thr Gly Met Thr Gly Gly Thr Arg Arg Glu Glu Ile Ala Gly Lys
65 70 75 80

Ile Asn Arg Thr Leu Ala Met Ala Ala Glu Glu Leu Asn Ile Pro Phe
85 90 95

Gly Val Gly Ser Gln Arg Ala Met Ile Glu Lys Pro Glu Thr Trp Glu
100 105 110

Ser Tyr Tyr Val Arg Asp Val Ala Pro Asp Ile Phe Leu Ile Gly Asn
115 120 125

Leu Gly Ala Pro Gln Phe Gly Lys Asn Ala Lys Lys Arg Tyr Ser Val
130 135 140

Lys Glu Val Leu Tyr Ala Ile Glu Lys Ile Glu Ala Asp Ala Ile Ala
145 150 155 160

Ile His Met Asn Pro Leu Gln Glu Ser Val Gln Pro Glu Gly Asp Thr
165 170 175

Thr Tyr Ala Gly Val Leu Glu Ala Leu Ala Glu Ile Lys Ser Ser Ile
 180 185 190

Asn Tyr Pro Val Ile Ala Lys Glu Thr Gly Ala Gly Val Ser Lys Glu

195	200	205
Val Ala Ile Glu Leu Glu Ser Val Gly Ile Asp Ala Ile Asp Ile Ser		
210	215	220
Gly Leu Gly Gly Thr Ser Trp Ser Ala Val Glu Tyr Tyr Arg Ala Lys		
225	230	235
Asp Ser Glu Lys Arg Lys Ile Ala Leu Lys Phe Trp Asp Trp Gly Ile		
245	250	255
Lys Thr Ala Ile Ser Leu Ala Glu Val Arg Trp Ala Thr Asn Leu Pro		
260	265	270
Ile Ile Ala Ser Gly Gly Met Arg Asp Gly Val Met Met Ala Lys Ala		
275	280	285
Leu Ala Met Gly Ala Ser Leu Val Gly Ile Ala Leu Pro Val Leu Arg		
290	295	300
Pro Ala Ala Arg Gly Asp Val Glu Gly Val Val Arg Ile Ile Arg Gly		
305	310	315
Tyr Ala Glu Glu Ile Lys Asn Val Met Phe Leu Val Gly Ala Arg Asn		
325	330	335
Ile Arg Glu Leu Arg Arg Val Pro Leu Val Ile Thr Gly Phe Val Arg		
340	345	350
Glu Trp Leu Leu Gln Arg Ile Asp Leu Asn Ser Tyr Leu Arg Ser Arg		
355	360	365
Phe Lys His		
370		
<210> 70		
<211> 349		
<212> PRT		
<213> Methanobacterium thermoautotrophicum		
<400> 70		
Met Ile Ser Asp Arg Lys Leu Glu His Leu Ile Leu Cys Ala Ser Cys		
1	5	10
15		
Asp Val Glu Tyr Arg Lys Lys Thr Gly Phe Glu Asp Ile Glu Ile Val		
20	25	30
His Arg Ala Ile Pro Glu Ile Asn Lys Glu Lys Ile Asp Ile Ser Leu		
35	40	45
Asp Phe Leu Gly Arg Glu Leu Ser Ser Pro Val Met Ile Ser Ala Ile		

50	55	60
Thr Gly Gly His Pro Ala Ser Met Lys Ile Asn Arg Glu Leu Ala Arg		
65	70	75
Ala Ala Glu Lys Leu Gly Ile Ala Leu Gly Leu Gly Ser Gln Arg Ala		
85	90	95
Gly Val Glu His Pro Glu Leu Glu Gly Thr Tyr Thr Ile Ala Arg Glu		
100	105	110
Glu Ala Pro Ser Ala Met Leu Ile Gly Asn Ile Gly Ser Ser His Ile		
115	120	125
Glu Tyr Ala Glu Arg Ala Val Glu Met Ile Asp Ala Asp Ala Leu Ala		
130	135	140
Val His Leu Asn Pro Leu Gln Glu Ser Ile Gln Pro Gly Gly Asp Val		
145	150	155
Asp Ser Ser Gly Ala Leu Glu Ser Ile Ser Ala Ile Val Glu Ser Val		
165	170	175
Asp Val Pro Val Met Val Lys Glu Thr Gly Ala Gly Ile Cys Ser Glu		
180	185	190
Asp Ala Ile Glu Leu Glu Ser Cys Gly Val Ser Ala Ile Asp Val Ala		
195	200	205
Gly Ala Gly Gly Thr Ser Trp Ala Ala Val Glu Thr Tyr Arg Ala Asp		
210	215	220
Asp Arg Tyr Leu Gly Glu Leu Phe Trp Asp Trp Gly Ile Pro Thr Ala		
225	230	235
Ala Ser Thr Val Glu Val Val Glu Ser Val Ser Ile Pro Val Ile Ala		
245	250	255
Ser Gly Gly Ile Arg Ser Gly Ile Asp Ala Ala Lys Ala Ile Ser Leu		
260	265	270
Gly Ala Glu Met Val Gly Ile Ala Leu Pro Val Leu Glu Ala Ala Gly		
275	280	285
His Gly Tyr Arg Glu Val Ile Lys Val Ile Glu Gly Phe Asn Glu Ala		
290	295	300
Leu Arg Thr Ala Met Tyr Leu Ala Gly Ala Glu Thr Leu Asp Asp Leu		
305	310	315
Lys Lys Ser Pro Val Ile Ile Thr Gly His Thr Gly Glu Trp Leu Asn		
325	330	335
Gln Arg Gly Phe Glu Thr Lys Lys Tyr Ala Arg Arg Ser		
340	345	

<210> 71

<211> 359

<212> PRT

<213> Methanococcus jannaschii

<400> 71

Met Val Asn Asn Arg Asn Glu Ile Glu Val Arg Lys Leu Glu His Ile
1 5 10 15

Phe Leu Cys Ser Tyr Cys Asn Val Glu Tyr Glu Lys Thr Thr Leu Leu
20 25 30

Glu Asp Ile Glu Leu Ile His Lys Gly Thr Cys Gly Ile Asn Phe Asn
35 40 45

Asp Ile Glu Thr Glu Ile Glu Leu Phe Gly Lys Lys Leu Ser Ala Pro
50 55 60

Ile Ile Val Ser Gly Met Thr Gly His Ser Lys Ala Lys Glu Ile
65 70 75 80

Asn Lys Asn Ile Ala Lys Ala Val Glu Glu Leu Gly Leu Gly Met Gly
85 90 95

Val Gly Ser Gln Arg Ala Ala Ile Val Asn Asp Glu Leu Ile Asp Thr
100 105 110

Tyr Ser Ile Val Arg Asp Tyr Thr Asn Asn Leu Val Ile Gly Asn Leu
115 120 125

Gly Ala Val Asn Phe Ile Val Asp Asp Trp Asp Glu Glu Ile Ile Asp
130 135 140

Lys Ala Ile Glu Met Ile Asp Ala Asp Ala Ile Ala Ile His Phe Asn
145 150 155 160

Pro Leu Gln Glu Ile Ile Gln Pro Glu Gly Asp Leu Asn Phe Lys Asn
165 170 175

Leu Tyr Lys Leu Lys Glu Ile Ile Ser Asn Tyr Lys Lys Ser Tyr Lys
180 185 190

Asn Ile Pro Phe Ile Ala Lys Gln Val Gly Glu Gly Phe Ser Lys Glu
195 200 205

Asp Ala Leu Ile Leu Lys Asp Ile Gly Phe Asp Ala Ile Asp Val Gln
210 215 220

Gly Ser Gly Gly Thr Ser Trp Ala Lys Val Glu Ile Tyr Arg Val Lys
225 230 235 240

Glu Glu Glu Ile Lys Arg Leu Ala Glu Lys Phe Ala Asn Trp Gly Ile

245	250	255
Pro Thr Ala Ala Ser Ile Phe Glu Val Lys Ser Val Tyr Asp Gly Ile		
260	265	270
Val Ile Gly Ser Gly Gly Ile Arg Gly Gly Leu Asp Ile Ala Lys Cys		
275	280	285
Ile Ala Ile Gly Cys Asp Cys Cys Ser Val Ala Leu Pro Ile Leu Lys		
290	295	300
Ala Ser Leu Lys Gly Trp Glu Glu Val Val Lys Val Leu Glu Ser Tyr		
305	310	315
Ile Lys Glu Leu Lys Ile Ala Met Phe Leu Val Gly Ala Glu Asn Ile		
325	330	335
Glu Glu Leu Lys Lys Thr Ser Tyr Ile Val Lys Gly Thr Leu Lys Glu		
340	345	350
Trp Ile Ser Gln Arg Leu Lys		
355		
<210> 72		
<211> 348		
<212> PRT		
<213> Thermoplasma acidophilum		
<400> 72		
Met Ile Gly Lys Arg Lys Glu Glu His Ile Arg Ile Ala Glu Asn Glu		
1	5	10
Asp Val Ser Ser Phe His Asn Phe Trp Asp Asp Ile Ser Leu Met His		
20	25	30
Glu Ala Asp Pro Glu Val Asn Tyr Asp Glu Ile Asp Thr Ser Val Asp		
35	40	45
Phe Leu Gly Lys Lys Leu Lys Phe Pro Met Ile Ile Ser Ser Met Thr		
50	55	60
Gly Gly Ala Glu Ile Ala Lys Asn Ile Asn Arg Asn Leu Ala Val Ala		
65	70	75
Ala Glu Arg Phe Gly Ile Gly Met Gly Val Gly Ser Met Arg Ala Ala		
85	90	95
Ile Val Asp Arg Ser Ile Glu Asp Thr Tyr Ser Val Ile Asn Glu Ser		
100	105	110
His Val Pro Leu Lys Ile Ala Asn Ile Gly Ala Pro Gln Leu Val Arg		

115	120	125
Gln Asp Lys Asp Ala Val Ser Asn Arg Asp Ile Ala Tyr Ile Tyr Asp		
130	135	140
Leu Ile Lys Ala Asp Phe Leu Ala Val His Phe Asn Phe Leu Gln Glu		
145	150	155
Met Val Gln Pro Glu Gly Asp Arg Asn Ser Lys Gly Val Ile Asp Arg		
165	170	175
Ile Lys Asp Leu Ser Gly Ser Phe Asn Ile Ile Ala Lys Glu Thr Gly		
180	185	190
Ser Gly Phe Ser Arg Arg Thr Ala Glu Arg Leu Ile Asp Ala Gly Val		
195	200	205
Lys Ala Ile Glu Val Ser Gly Val Ser Gly Thr Thr Phe Ala Ala Val		
210	215	220
Glu Tyr Tyr Arg Ala Arg Lys Glu Asn Asn Leu Glu Lys Met Arg Ile		
225	230	235
Gly Glu Thr Phe Trp Asn Trp Gly Ile Pro Ser Pro Ala Ser Val Tyr		
245	250	255
Tyr Cys Ser Asp Leu Ala Pro Val Ile Gly Ser Gly Gly Leu Arg Asn		
260	265	270
Gly Leu Asp Leu Ala Lys Ala Ile Ala Met Gly Ala Thr Ala Gly Gly		
275	280	285
Phe Ala Arg Ser Leu Leu Lys Asp Ala Asp Thr Asp Pro Glu Met Leu		
290	295	300
Met Lys Asn Ile Glu Leu Ile Gln Arg Glu Phe Arg Val Ala Leu Phe		
305	310	315
Leu Thr Gly Asn Lys Asn Val Tyr Glu Leu Lys Phe Thr Lys Lys Val		
325	330	335
Ile Val Asp Pro Leu Arg Ser Trp Leu Glu Ala Lys		
340	345	

<210> 73

<211> 357

<212> PRT

<213> Leishmania major

<400> 73

Met Ser Ser Arg Asp Cys Thr Val Asp Arg Glu Ala Ala Val Gln Lys

1	5	10	15
Arg Lys Lys Asp His Ile Asp Ile Cys Leu His Gln Asp Val Glu Pro			
20	25		30
His Lys Arg Arg Thr Ser Ile Trp Asn Lys Tyr Thr Leu Pro Tyr Lys			
35	40		45
Ala Leu Pro Glu Val Asp Leu Gln Lys Ile Asp Thr Ser Cys Glu Phe			
50	55	60	
Met Gly Lys Arg Ile Ser Phe Pro Phe Phe Ile Ser Ser Met Thr Gly			
65	70	75	80
Gly Glu Ala His Gly Arg Val Ile Asn Glu Asn Leu Ala Lys Ala Cys			
85	90	95	
Glu Ala Glu Lys Ile Pro Phe Gly Leu Gly Ser Met Arg Ile Ile Asn			
100	105	110	
Arg Tyr Ala Ser Ala Val His Thr Phe Asn Val Lys Glu Phe Cys Pro			
115	120	125	
Ser Val Pro Met Leu Ala Asn Ile Gly Leu Val Gln Leu Asn Tyr Gly			
130	135	140	
Phe Gly Pro Lys Glu Val Asn Asn Leu Val Asn Ser Val Arg Ala Asp			
145	150	155	160
Gly Leu Cys Ile His Leu Asn His Thr Gln Glu Val Cys Gln Pro Glu			
165	170	175	
Gly Asp Thr Asn Phe Glu Gly Leu Ile Glu Lys Leu Arg Gln Leu Leu			
180	185	190	
Pro His Ile Lys Val Pro Val Leu Val Lys Gly Val Gly His Gly Ile			
195	200	205	
Asp Tyr Glu Ser Met Val Ala Ile Lys Ala Ser Gly Val Lys Tyr Val			
210	215	220	
Asp Val Ser Gly Cys Gly Gly Thr Ser Trp Ala Trp Ile Glu Gly Arg			
225	230	235	240
Arg Gln Pro Tyr Lys Ala Glu Glu Asn Ile Gly Tyr Leu Leu Arg			
245	250	255	
Asp Ile Gly Val Pro Thr Asp Val Cys Leu Arg Glu Ser Ala Pro Leu			
260	265	270	
Thr Val Asn Gly Asp Leu His Leu Ile Ala Gly Gly Ile Arg Asn			
275	280	285	
Gly Met Asp Val Ala Lys Ala Leu Met Met Gly Ala Glu Tyr Ala Thr			
290	295	300	
Ala Ala Met Pro Phe Leu Ala Ala Leu Glu Ser Ser Glu Ala Val			

305	310	315	320
Arg Ala Val Ile Gln Arg Met Arg Gln Glu Leu Arg Val Ser Met Phe			
325		330	335
Thr Cys Gly Ala Arg Asn Ile Glu Glu Leu Arg Arg Met Lys Val Ile			
340	345		350
Glu Leu Gly His Leu			
355			
 <210> 74			
 <211> 398			
 <212> PRT			
 <213> Streptococcus pneumoniae			
()			
 <400> 74			
Met Asn Asp Lys Thr Glu Val Asn Met Thr Ile Gly Ile Asp Lys Ile			
1	5	10	15
Gly Phe Ala Thr Ser Gln Tyr Val Leu Lys Leu Gln Asp Leu Ala Glu			
20	25		30
Ala Arg Gly Ile Asp Pro Glu Lys Leu Ser Lys Gly Leu Leu Leu Lys			
35	40		45
Glu Leu Ser Ile Ala Pro Leu Thr Glu Asp Ile Val Thr Leu Ala Ala			
50	55	60	
Ser Ala Ser Asp Ser Ile Leu Thr Glu Gln Glu Arg Gln Glu Val Asp			
65	70	75	80
Met Val Ile Val Ala Thr Glu Ser Gly Ile Asp Gln Ser Lys Ala Ala			
85	90		95
Ala Val Phe Val His Gly Leu Leu Gly Ile Gln Pro Phe Ala Arg Ser			
100	105		110
Phe Glu Ile Lys Glu Ala Cys Tyr Gly Ala Thr Ala Ala Leu His Tyr			
115	120	125	
Ala Lys Leu His Val Glu Asn Ser Pro Glu Ser Lys Val Leu Val Ile			
130	135	140	
Ala Ser Asp Ile Ala Lys Tyr Gly Ile Glu Thr Pro Gly Glu Pro Thr			
145	150	155	160
Gln Gly Ala Gly Ser Val Ala Met Leu Ile Thr Gln Asn Pro Arg Met			
165	170		175
Met Ala Phe Asn Asn Asp Asn Val Ala Gln Thr Arg Asp Ile Met Asp			
()			

180	185	190
Phe Trp Arg Pro Asn Tyr Ser Thr	Thr Pro Tyr Val Asn Gly Val Tyr	
195	200	205
Ser Thr Gln Gln Tyr Leu Asp Ser	Leu Lys Thr Thr Trp Leu Glu Tyr	
210	215	220
Gln Lys Arg Tyr Gln Leu Thr Leu Asp Asp	Phe Ala Ala Val Cys Phe	
225	230	235
His Leu Pro Tyr Pro Lys Leu Ala Leu	Lys Gly Leu Lys Lys Ile Met	
245	250	255
Asp Lys Asn Leu Pro Gln Glu Lys Lys Asp	Leu Leu Gln Lys His Phe	
260	265	270
Asp Gln Ser Ile Leu Tyr Ser Gln Lys Val	Gly Asn Ile Tyr Thr Gly	
275	280	285
Ser Leu Phe Leu Gly Leu Leu Ser	Leu Leu Glu Asn Thr Asp Ser Leu	
290	295	300
Lys Ala Gly Asp Lys Ile Ala Leu Tyr Ser	Tyr Gly Ser Gly Ala Val	
305	310	315
320		
Ala Glu Phe Phe Ser Gly Glu Leu Val	Glu Gly Tyr Glu Ala Tyr Leu	
325	330	335
Asp Lys Asp Arg Leu Asn Lys Leu Asn Gln Arg	Thr Ala Leu Ser Val	
340	345	350
Ala Asp Tyr Glu Lys Val Phe Phe Glu Glu Val	Asn Leu Asp Glu Thr	
355	360	365
Asn Ser Ala Gln Phe Ala Gly Tyr Glu Asn Gln Asp	Phe Ala Leu Val	
370	375	380
Glu Ile Leu Asp His Gln Arg Arg Tyr Ser	Lys Val Glu Lys	
385	390	395
<210> 75		
<211> 391		
<212> PRT		
<213> Streptococcus pyogenes		
<400> 75		
Met Thr Ile Gly Ile Asp Lys Ile Gly Phe Ala Thr Ser Gln Tyr Val		
1	5	10
15		
Leu Lys Leu Glu Asp Leu Ala Leu Ala Arg Gln Val Asp Pro Ala Lys		

20	25	30
Phe Ser Gln Gly Leu Leu Ile Glu Ser Phe Ser Val Ala Pro Ile Thr		
35	40	45
Glu Asp Ile Ile Thr Leu Ala Ala Ser Ala Ala Asp Gln Ile Leu Thr		
50	55	60
Asp Glu Asp Arg Ala Lys Ile Asp Met Val Ile Leu Ala Thr Glu Ser		
65	70	80
Ser Thr Asp Gln Ser Lys Ala Ser Ala Ile Tyr Val His His Leu Val		
85	90	95
Gly Ile Gln Pro Phe Ala Arg Ser Phe Glu Val Lys Gln Ala Cys Tyr		
100	105	110
Ser Ala Thr Ala Ala Leu Asp Tyr Ala Lys Leu His Val Ala Ser Lys		
115	120	125
Pro Asp Ser Arg Val Leu Val Ile Ala Ser Asp Ile Ala Arg Tyr Gly		
130	135	140
Val Gly Ser Pro Gly Glu Ser Thr Gln Gly Ser Gly Ser Ile Ala Leu		
145	150	155
Leu Val Thr Ala Asp Pro Arg Ile Leu Ala Leu Asn Glu Asp Asn Val		
165	170	175
Ala Gln Thr Arg Asp Ile Met Asp Phe Trp Arg Pro Asn Tyr Ser Phe		
180	185	190
Thr Pro Tyr Val Asp Gly Ile Tyr Ser Thr Lys Gln Tyr Leu Asn Cys		
195	200	205
Leu Glu Thr Thr Trp Gln Ala Tyr Gln Lys Arg Glu Asn Leu Gln Leu		
210	215	220
Ser Asp Phe Ala Ala Val Cys Phe His Ile Pro Phe Pro Lys Leu Ala		
225	230	240
Leu Lys Gly Leu Asn Asn Ile Met Asp Asn Thr Val Pro Pro Glu His		
245	250	255
Arg Glu Lys Leu Ile Glu Ala Phe Gln Ala Ser Ile Thr Tyr Ser Lys		
260	265	270
Gln Ile Gly Asn Ile Tyr Thr Gly Ser Leu Tyr Leu Gly Leu Leu Ser		
275	280	285
Leu Leu Glu Asn Ser Lys Val Leu Gln Ser Gly Asp Lys Ile Gly Phe		
290	295	300
Phe Ser Tyr Gly Ser Gly Ala Val Ser Glu Phe Tyr Ser Gly Gln Leu		
305	310	320
Val Ala Gly Tyr Asp Lys Met Leu Met Thr Asn Arg Gln Ala Leu Leu		

325

330

335

Asp Gln Arg Thr Arg Leu Ser Val Ser Lys Tyr Glu Asp Leu Phe Tyr
 340 345 350

Glu Gln Val Gln Leu Asp Asp Asn Gly Asn Ala Asn Phe Asp Ile Tyr
 355 360 365

Leu Thr Gly Lys Phe Ala Leu Thr Ala Ile Lys Glu His Gln Arg Ile
 370 375 380

Tyr His Thr Asn Asp Lys Asn
 385 390

<210> 76

<211> 383

<212> PRT

<213> Enterococcus faecalis

<400> 76

Met Thr Ile Gly Ile Asp Lys Ile Ser Phe Phe Val Pro Pro Tyr Tyr
 1 5 10 15

Ile Asp Met Thr Ala Leu Ala Glu Ala Arg Asn Val Asp Pro Gly Lys
 20 25 30

Phe His Ile Gly Ile Gly Gln Asp Gln Met Ala Val Asn Pro Ile Ser
 35 40 45

Gln Asp Ile Val Thr Phe Ala Ala Asn Ala Ala Glu Ala Ile Leu Thr
 50 55 60

Lys Glu Asp Lys Glu Ala Ile Asp Met Val Ile Val Gly Thr Glu Ser
 65 70 75 80

Ser Ile Asp Glu Ser Lys Ala Ala Val Val Leu His Arg Leu Met
 85 90 95

Gly Ile Gln Pro Phe Ala Arg Ser Phe Glu Ile Lys Glu Ala Cys Tyr
 100 105 110

Gly Ala Thr Ala Gly Leu Gln Leu Ala Lys Asn His Val Ala Leu His
 115 120 125

Pro Asp Lys Lys Val Leu Val Val Ala Ala Asp Ile Ala Lys Tyr Gly
 130 135 140

Leu Asn Ser Gly Gly Glu Pro Thr Gln Gly Ala Gly Ala Val Ala Met
 145 150 155 160

Leu Val Ala Ser Glu Pro Arg Ile Leu Ala Leu Lys Glu Asp Asn Val

165	170	175
Met Leu Thr Gln Asp Ile Tyr Asp Phe Trp Arg Pro Thr Gly His Pro 180	185	190
Tyr Pro Met Val Asp Gly Pro Leu Ser Asn Glu Thr Tyr Ile Gln Ser 195	200	205
Phe Ala Gln Val Trp Asp Glu His Lys Lys Arg Thr Gly Leu Asp Phe 210	215	220
Ala Asp Tyr Asp Ala Leu Ala Phe His Ile Pro Tyr Thr Lys Met Gly 225	230	235
Lys Lys Ala Leu Leu Ala Lys Ile Ser Asp Gln Thr Glu Ala Glu Gln 245	250	255
Glu Arg Ile Leu Ala Arg Tyr Glu Glu Ser Ile Ile Tyr Ser Arg Arg 260	265	270
Val Gly Asn Leu Tyr Thr Gly Ser Leu Tyr Leu Gly Leu Ile Ser Leu 275	280	285
Leu Glu Asn Ala Thr Thr Leu Thr Ala Gly Asn Gln Ile Gly Leu Phe 290	295	300
Ser Tyr Gly Ser Gly Ala Val Ala Glu Phe Phe Thr Gly Glu Leu Val 305	310	315
Ala Gly Tyr Gln Asn His Leu Gln Lys Glu Thr His Leu Ala Leu Leu 325	330	335
Asp Asn Arg Thr Glu Leu Ser Ile Ala Glu Tyr Glu Ala Met Phe Ala 340	345	350
Glu Thr Leu Asp Thr Asp Ile Asp Gln Thr Leu Glu Asp Glu Leu Lys 355	360	365
Tyr Ser Ile Ser Ala Ile Asn Asn Thr Val Arg Ser Tyr Arg Asn 370	375	380
<210> 77		
<211> 384		
<212> PRT		
<213> Enterococcus faecium		
<400> 77		
Met Lys Ile Gly Ile Asp Arg Leu Ser Phe Phe Ile Pro Asn Leu Tyr 1 5 10 15		
Leu Asp Met Thr Glu Leu Ala Glu Ser Arg Gly Asp Asp Pro Ala Lys		

20	25	30	
Tyr His Ile Gly Ile Gly Gln Asp Gln Met Ala Val Asn Arg Ala Asn			
35	40	45	
Glu Asp Ile Ile Thr Leu Gly Ala Asn Ala Ala Ser Lys Ile Val Thr			
50	55	60	
Glu Lys Asp Arg Glu Leu Ile Asp Met Val Ile Val Gly Thr Glu Ser			
65	70	75	80
Gly Ile Asp His Ser Lys Ala Ser Ala Val Ile Ile His His Leu Leu			
85	90	95	
Lys Ile Gln Ser Phe Ala Arg Ser Phe Glu Val Lys Glu Ala Cys Tyr			
100	105	110	
Gly Gly Thr Ala Ala Leu His Met Ala Lys Glu Tyr Val Lys Asn His			
115	120	125	
Pro Glu Arg Lys Val Leu Val Ile Ala Ser Asp Ile Ala Arg Tyr Gly			
130	135	140	
Leu Ala Ser Gly Gly Glu Val Thr Gln Gly Val Gly Ala Val Ala Met			
145	150	155	160
Met Ile Thr Gln Asn Pro Arg Ile Leu Ser Ile Glu Asp Asp Ser Val			
165	170	175	
Phe Leu Thr Glu Asp Ile Tyr Asp Phe Trp Arg Pro Asp Tyr Ser Glu			
180	185	190	
Phe Pro Val Val Asp Gly Pro Leu Ser Asn Ser Thr Tyr Ile Glu Ser			
195	200	205	
Phe Gln Lys Val Trp Asn Arg His Lys Glu Leu Ser Gly Arg Gly Leu			
210	215	220	
Glu Asp Tyr Gln Ala Ile Ala Phe His Ile Pro Tyr Thr Lys Met Gly			
225	230	235	240
Lys Lys Ala Leu Gln Ser Val Leu Asp Gln Thr Asp Glu Asp Asn Gln			
245	250	255	
Glu Arg Leu Met Ala Arg Tyr Glu Glu Ser Ile Arg Tyr Ser Arg Arg			
260	265	270	
Ile Gly Asn Leu Tyr Thr Gly Ser Leu Tyr Leu Gly Leu Thr Ser Leu			
275	280	285	
Leu Glu Asn Ser Lys Ser Leu Gln Pro Gly Asp Arg Ile Gly Leu Phe			
290	295	300	
Ser Tyr Gly Ser Gly Ala Val Ser Glu Phe Phe Thr Gly Tyr Leu Glu			
305	310	315	320
Glu Asn Tyr Gln Glu Tyr Leu Phe Ala Gln Ser His Gln Glu Met Leu			

	325	330	335
Asp Ser Arg Thr Arg Ile Thr Val Asp Glu Tyr Glu Thr Ile Phe Ser			
340	345		350
Glu Thr Leu Pro Glu His Gly Glu Cys Ala Glu Tyr Thr Ser Asp Val			
355	360		365
Pro Phe Ser Ile Thr Lys Ile Glu Asn Asp Ile Arg Tyr Tyr Lys Ile			
370	375		380
<210> 78			
<211> 388			
<212> PRT			
<213> Staphylococcus haemolyticus			
<400> 78			
Met Ser Ile Gly Ile Asp Lys Ile Asn Phe Tyr Val Pro Lys Tyr Tyr			
1	5	10	15
Val Asp Met Ala Lys Leu Ala Glu Ala Arg Gln Val Asp Pro Asn Lys			
20	25		30
Phe Leu Ile Gly Ile Gly Gln Thr Gln Met Ala Val Ser Pro Val Ser			
35	40		45
Gln Asp Ile Val Ser Met Gly Ala Asn Ala Ala Lys Asp Ile Ile Thr			
50	55	60	
Asp Asp Asp Lys Lys His Ile Gly Met Val Ile Val Ala Thr Glu Ser			
65	70	75	80
Ala Ile Asp Asn Ala Lys Ala Ala Ala Val Gln Ile His Asn Leu Leu			
85	90		95
Gly Val Gln Pro Phe Ala Arg Cys Phe Glu Met Lys Glu Ala Cys Tyr			
100	105		110
Ala Ala Thr Pro Ala Ile Gln Leu Ala Lys Asp Tyr Ile Glu Lys Arg			
115	120		125
Pro Asn Glu Lys Val Leu Val Ile Ala Ser Asp Thr Ala Arg Tyr Gly			
130	135		140
Ile Gln Ser Gly Gly Glu Pro Thr Gln Gly Ala Gly Ala Val Ala Met			
145	150		155
Leu Ile Ser Asn Asn Pro Ser Ile Leu Glu Leu Asn Asp Asp Ala Val			
165	170		175
Ala Tyr Thr Glu Asp Val Tyr Asp Phe Trp Arg Pro Thr Gly His Lys			

180	185	190
Tyr Pro Leu Val Ala Gly Ala Leu Ser Lys Asp Ala Tyr Ile Lys Ser		
195	200	205
Phe Gln Glu Ser Trp Asn Glu Tyr Ala Arg Arg Glu Asp Lys Thr Leu		
210	215	220
Ser Asp Phe Glu Ser Leu Cys Phe His Val Pro Phe Thr Lys Met Gly		
225	230	235
Lys Lys Ala Leu Asp Ser Ile Ile Asn Asp Ala Asp Glu Thr Thr Gln		
245	250	255
Glu Arg Leu Thr Ser Gly Tyr Glu Asp Ala Val Tyr Tyr Asn Arg Tyr		
260	265	270
Val Gly Asn Ile Tyr Thr Gly Ser Leu Tyr Leu Ser Leu Ile Ser Leu		
275	280	285
Leu Glu Asn Arg Ser Leu Lys Gly Gly Gln Thr Ile Gly Leu Phe Ser		
290	295	300
Tyr Gly Ser Gly Ser Val Gly Glu Phe Phe Ser Ala Thr Leu Val Glu		
305	310	315
Gly Tyr Glu Lys Gln Leu Asp Ile Glu Gly His Lys Ala Leu Leu Asn		
325	330	335
Glu Arg Gln Glu Val Ser Val Glu Asp Tyr Glu Ser Phe Phe Lys Arg		
340	345	350
Phe Asp Asp Leu Glu Phe Asp His Ala Thr Glu Gln Thr Asp Asp Asp		
355	360	365
Lys Ser Ile Tyr Tyr Leu Glu Asn Ile Gln Asp Asp Ile Arg Gln Tyr		
370	375	380
His Ile Pro Lys		
385		
<210> 79		
<211> 388		
<212> PRT		
<213> <i>Staphylococcus epidermidis</i>		
<400> 79		
Met Asn Ile Gly Ile Asp Lys Ile Ser Phe Tyr Val Pro Lys Tyr Tyr		
1	5	10
15		
Val Asp Met Ala Lys Leu Ala Glu Ala Arg Gln Val Asp Pro Asn Lys		

20

25

30

Phe	Leu	Ile	Gly	Ile	Gly	Gln	Thr	Glu	Met	Thr	Val	Ser	Pro	Val	Asn
35						40						45			
Gln	Asp	Ile	Val	Ser	Met	Gly	Ala	Asn	Ala	Ala	Lys	Asp	Ile	Ile	Thr
50						55					60				
Glu	Glu	Asp	Lys	Lys	Asn	Ile	Gly	Met	Val	Ile	Val	Ala	Thr	Glu	Ser
65						70				75			80		
Ala	Ile	Asp	Asn	Ala	Lys	Ala	Ala	Ala	Val	Gln	Ile	His	His	Leu	Leu
						85				90			95		
Gly	Ile	Gln	Pro	Phe	Ala	Arg	Cys	Phe	Glu	Met	Lys	Glu	Ala	Cys	Tyr
						100				105			110		
Ala	Ala	Thr	Pro	Ala	Ile	Gln	Leu	Ala	Lys	Asp	Tyr	Leu	Ala	Gln	Arg
						115				120			125		
Pro	Asn	Glu	Lys	Val	Leu	Val	Ile	Ala	Ser	Asp	Thr	Ala	Arg	Tyr	Gly
						130				135			140		
Ile	His	Ser	Gly	Gly	Glu	Pro	Thr	Gln	Gly	Ala	Gly	Ala	Val	Ala	Met
145						150				155			160		
Met	Ile	Ser	His	Asp	Pro	Ser	Ile	Leu	Lys	Leu	Asn	Asp	Asp	Ala	Val
						165				170			175		
Ala	Tyr	Thr	Glu	Asp	Val	Tyr	Asp	Phe	Trp	Arg	Pro	Thr	Gly	His	Gln
						180				185			190		
Tyr	Pro	Leu	Val	Ala	Gly	Ala	Leu	Ser	Lys	Asp	Ala	Tyr	Ile	Lys	Ser
						195				200			205		
Phe	Gln	Glu	Ser	Trp	Asn	Glu	Tyr	Ala	Arg	Arg	His	Asn	Lys	Thr	Leu
						210				215			220		
Ala	Asp	Phe	Ala	Ser	Leu	Cys	Phe	His	Val	Pro	Phe	Thr	Lys	Met	Gly
225						230				235			240		
Gln	Lys	Ala	Leu	Asp	Ser	Ile	Ile	Asn	His	Ala	Asp	Glu	Thr	Thr	Gln
						245				250			255		
Asp	Arg	Leu	Asn	Ser	Ser	Tyr	Gln	Asp	Ala	Val	Asp	Tyr	Asn	Arg	Tyr
						260				265			270		
Val	Gly	Asn	Ile	Tyr	Thr	Gly	Ser	Leu	Tyr	Leu	Ser	Leu	Ile	Ser	Leu
						275				280			285		
Leu	Glu	Thr	Arg	Asp	Leu	Lys	Gly	Gly	Gln	Thr	Ile	Gly	Leu	Phe	Ser
						290				295			300		
Tyr	Gly	Ser	Gly	Ser	Val	Gly	Glu	Phe	Phe	Ser	Gly	Thr	Leu	Val	Asp
305						310				315			320		
Gly	Phe	Lys	Glu	Gln	Leu	Asp	Val	Glu	Arg	His	Lys	Ser	Leu	Leu	Asn

325

330

335

Asn Arg Ile Glu Val Ser Val Asp Glu Tyr Glu His Phe Phe Lys Arg
340 345 350

Phe Asp Gln Leu Glu Leu Asn His Glu Leu Glu Lys Ser Asn Ala Asp
355 360 365

Arg Asp Ile Phe Tyr Leu Lys Ser Ile Asp Asn Asn Ile Arg Glu Tyr
370 375 380

His Ile Ala Glu
385

<210> 80

<211> 388

<212> PRT

<213> Staphylococcus aureus

<400> 80

Met Thr Ile Gly Ile Asp Lys Ile Asn Phe Tyr Val Pro Lys Tyr Tyr
1 5 10 15

Val Asp Met Ala Lys Leu Ala Glu Ala Arg Gln Val Asp Pro Asn Lys
20 25 30

Phe Leu Ile Gly Ile Gly Gln Thr Glu Met Ala Val Ser Pro Val Asn
35 40 45

Gln Asp Ile Val Ser Met Gly Ala Asn Ala Ala Lys Asp Ile Ile Thr
50 55 60

Asp Glu Asp Lys Lys Ile Gly Met Val Ile Val Ala Thr Glu Ser
65 70 75 80

Ala Val Asp Ala Ala Lys Ala Ala Val Gln Ile His Asn Leu Leu
85 90 95

Gly Ile Gln Pro Phe Ala Arg Cys Phe Glu Met Lys Glu Ala Cys Tyr
100 105 110

Ala Ala Thr Pro Ala Ile Gln Leu Ala Lys Asp Tyr Leu Ala Thr Arg
115 120 125

Pro Asn Glu Lys Val Leu Val Ile Ala Thr Asp Thr Ala Arg Tyr Gly
130 135 140

Leu Asn Ser Gly Gly Glu Pro Thr Gln Gly Ala Gly Ala Val Ala Met
145 150 155 160

Val Ile Ala His Asn Pro Ser Ile Leu Ala Leu Asn Glu Asp Ala Val

165	170	175
Ala Tyr Thr Glu Asp Val Tyr Asp Phe Trp Arg Pro Thr Gly His Lys 180	185	190
Tyr Pro Leu Val Asp Gly Ala Leu Ser Lys Asp Ala Tyr Ile Arg Ser 195	200	205
Phe Gln Gln Ser Trp Asn Glu Tyr Ala Lys Arg Gln Gly Lys Ser Leu 210	215	220
Ala Asp Phe Ala Ser Leu Cys Phe His Val Pro Phe Thr Lys Met Gly 225	230	235
Lys Lys Ala Leu Glu Ser Ile Ile Asp Asn Ala Asp Glu Thr Thr Gln 245	250	255
Glu Arg Leu Arg Ser Gly Tyr Glu Asp Ala Val Asp Tyr Asn Arg Tyr 260	265	270
Val Gly Asn Ile Tyr Thr Gly Ser Leu Tyr Leu Ser Leu Ile Ser Leu 275	280	285
Leu Glu Asn Arg Asp Leu Gln Ala Gly Glu Thr Ile Gly Leu Phe Ser 290	295	300
Tyr Gly Ser Gly Ser Val Val Glu Phe Tyr Ser Ala Thr Leu Val Val 305	310	315
Gly Tyr Lys Asp His Leu Asp Gln Ala Ala His Lys Ala Leu Leu Asn 325	330	335
Asn Arg Thr Glu Val Ser Val Asp Ala Tyr Glu Thr Phe Phe Lys Arg 340	345	350
Phe Asp Asp Val Glu Phe Asp Glu Gln Asp Ala Val His Glu Asp 355	360	365
Arg His Ile Phe Tyr Leu Ser Asn Ile Glu Asn Asn Val Arg Glu Tyr 370	375	380
His Arg Pro Glu 385		
<210> 81		
<211> 389		
<212> PRT		
<213> Staphylococcus carnosus		
<400> 81		
Met Thr Ile Gly Ile Asp Gln Leu Asn Phe Tyr Ile Pro Asn Phe Tyr		

1	5	10	15
Val Asp Met Ala Glu Leu Ala Glu Ala Arg Gly Val Asp Pro Asn Lys			
20	25		30
Phe Leu Ile Gly Ile Gly Gln Ser Gln Met Ala Val Ser Pro Val Ser			
35	40	45	
Gln Asp Ile Val Ser Met Gly Ala Asn Ala Ala Gln Pro Ile Leu Ser			
50	55	60	
Glu Gln Asp Lys Lys Asp Ile Thr Met Val Ile Val Ala Thr Glu Ser			
65	70	75	80
Ala Ile Asp Ser Ala Lys Ala Ser Ala Val Gln Ile His His Leu Leu			
85	90		95
Gly Ile Gln Pro Phe Ala Arg Cys Phe Glu Met Lys Glu Ala Cys Tyr			
100	105		110
Ala Ala Thr Pro Ala Ile Gln Leu Ala Lys Asp Tyr Leu Val Pro Arg			
115	120		125
Pro Lys Glu Lys Val Leu Val Ile Ala Ser Asp Thr Ala Arg Tyr Gly			
130	135		140
Leu Asn Ser Gly Gly Glu Pro Thr Gln Gly Ala Gly Ala Val Ala Met			
45	150	155	160
Al Ile Ser His Asn Pro Ser Ile Leu Glu Leu His Asp Asp Ser Val			
165	170		175
Ile Tyr Thr Glu Asp Val Tyr Asp Phe Trp Arg Pro Ser Gly Glu Ile			
180	185		190
Tyr Pro Leu Val Ala Gly Lys Leu Ser Lys Asp Ala Tyr Ile Lys Ser			
195	200		205
Le Gln Glu Ser Trp Asn Glu Tyr Ala Lys Arg His His Lys Ser Leu			
210	215		220
er Asp Phe Ala Ala Leu Cys Phe His Val Pro Phe Thr Lys Met Gly			
225	230		240
In Lys Ala Leu Asp Ser Ile Leu Thr Asp Ser Ala Ser Glu Asp Thr			
245	250		255
In Ala Arg Leu Asn Glu Gly Tyr Lys Ser Ala Thr Asp Tyr Asn Arg			
260	265		270
r Val Gly Asn Val Tyr Thr Gly Ser Leu Tyr Leu Ser Leu Ile Ser			
275	280		285
u Leu Glu Asn His Lys Leu Asn Gly Gly Asp Asn Ile Gly Leu Phe			
290	295		300
r Tyr Gly Ser Gly Ser Val Gly Glu Phe Phe Ser Ala Thr Leu Val			

305 310 315 320
Asp Asn Tyr Gln Asp His Leu Asp Val Lys Ala His Lys Ala Met Leu
325 330 335
Asp Asn Arg Lys Ala Leu Ser Val Glu Glu Tyr Glu Lys Phe Phe Asn
340 345 350
Arg Phe Asp Asn Leu Glu Phe Asp Thr Glu Thr Glu Leu Glu Val Glu
355 360 365
Pro Lys Gly Asn Phe Tyr Leu Lys Glu Ile Ser Asp Asn Ile Arg Tyr
370 375 380
Tyr Asp Thr Val Lys
385
<210> 82
<211> 389
<212> PRT
<213> Streptomyces sp. CL190

<400> 82
Met Ser Ile Ser Ile Gly Ile His Asp Leu Ser Phe Ala Thr Thr Glu
1 5 10 15
Phe Val Leu Pro His Thr Ala Leu Ala Glu Tyr Asn Gly Thr Glu Ile
20 25 30
Gly Lys Tyr His Val Gly Ile Gly Gln Gln Ser Met Ser Val Pro Ala
35 40 45
Ala Asp Glu Asp Ile Val Thr Met Ala Ala Thr Ala Ala Arg Pro Ile
50 55 60
Ile Glu Arg Asn Gly Lys Ser Arg Ile Arg Thr Val Val Phe Ala Thr
65 70 75 80
Glu Ser Ser Ile Asp Gln Ala Lys Ala Gly Gly Val Tyr Val His Ser
85 90 95
Leu Leu Gly Leu Glu Ser Ala Cys Arg Val Val Glu Leu Lys Gln Ala
100 105 110
Cys Tyr Gly Ala Thr Ala Ala Leu Gln Phe Ala Ile Gly Leu Val Arg
115 120 125
Arg Asp Pro Ala Gln Gln Val Leu Val Ile Ala Ser Asp Val Ser Lys
130 135 140
Tyr Glu Leu Asp Ser Pro Gly Glu Ala Thr Gln Gly Ala Ala Ala Val

145	150	155	160
Ala Met Leu Val Gly Ala Asp Pro Ala Leu Leu Arg Ile Glu Glu Pro			
165	170	175	
Ser Gly Leu Phe Thr Ala Asp Val Met Asp Phe Trp Arg Pro Asn Tyr			
180	185	190	
Leu Thr Thr Ala Leu Val Asp Gly Gln Glu Ser Ile Asn Ala Tyr Leu			
195	200	205	
Gln Ala Val Glu Gly Ala Trp Lys Asp Tyr Ala Glu Gln Asp Gly Arg			
210	215	220	
Ser Leu Glu Glu Phe Ala Ala Phe Val Tyr His Gln Pro Phe Thr Lys			
225	230	235	240
Met Ala Tyr Lys Ala His Arg His Leu Leu Asn Phe Asn Gly Tyr Asp			
245	250	255	
Thr Asp Lys Asp Ala Ile Glu Gly Ala Leu Gly Gln Thr Thr Ala Tyr			
260	265	270	
Asn Asn Val Ile Gly Asn Ser Tyr Thr Ala Ser Val Tyr Leu Gly Leu			
275	280	285	
Ala Ala Leu Leu Asp Gln Ala Asp Asp Leu Thr Gly Arg Ser Ile Gly			
290	295	300	
Phe Leu Ser Tyr Gly Ser Gly Ser Val Ala Glu Phe Phe Ser Gly Thr			
305	310	315	320
Val Val Ala Gly Tyr Arg Glu Arg Leu Arg Thr Glu Ala Asn Gln Glu			
325	330	335	
Ala Ile Ala Arg Arg Lys Ser Val Asp Tyr Ala Thr Tyr Arg Glu Leu			
340	345	350	
His Glu Tyr Thr Leu Pro Ser Asp Gly Gly Asp His Ala Thr Pro Val			
355	360	365	
Gln Thr Thr Gly Pro Phe Arg Leu Ala Gly Ile Asn Asp His Lys Arg			
370	375	380	
Ile Tyr Glu Ala Arg			
385			
<210> 83			
<211> 389			
<212> PRT			
<213> Streptomyces griseolosporeus			

<400> 83

Met Pro Leu Ala Ile Gly Ile His Asp Leu Ser Phe Ala Thr Gly Glu
1 5 10 15

Phe Val Leu Pro His Thr Ala Leu Ala Ala His Asn Gly Thr Glu Ile
20 25 30

Gly Lys Tyr His Ala Gly Ile Gly Gln Glu Ser Met Ser Val Pro Ala
35 40 45

Ala Asp Glu Asp Ile Val Thr Leu Ala Ala Thr Ala Ala Ala Pro Ile
50 55 60

Val Ala Arg His Gly Ser Asp Arg Ile Arg Thr Val Val Leu Ala Thr
65 70 75 80

Glu Ser Ser Ile Asp Gln Ala Lys Ser Ala Gly Val Tyr Val His Ser
85 90 95

Leu Leu Gly Leu Pro Ser Ala Thr Arg Val Val Glu Leu Lys Gln Ala
100 105 110

Cys Tyr Gly Ala Thr Ala Gly Leu Gln Phe Ala Ile Gly Leu Val Gln
115 120 125

Arg Asp Pro Ala Gln Gln Val Leu Val Ile Ala Ser Asp Val Ser Lys
130 135 140

Tyr Asp Leu Asp Ser Pro Gly Glu Ala Thr Gln Gly Ala Ala Ala Val
145 150 155 160

Ala Met Leu Val Gly Ala Asp Pro Gly Leu Val Arg Ile Glu Asp Pro
165 170 175

Ser Gly Leu Phe Thr Val Asp Val Met Asp Phe Trp Arg Pro Asn Tyr
180 185 190

Arg Thr Thr Ala Leu Val Asp Gly Gln Glu Ser Ile Gly Ala Tyr Leu
195 200 205

Gln Ala Val Glu Gly Ala Trp Lys Asp Tyr Ser Glu Arg Gly Gly His
210 215 220

Ser Leu Glu Gln Phe Ala Ala Phe Cys Tyr His Gln Pro Phe Thr Lys
225 230 235 240

Met Ala His Lys Ala His Arg His Leu Leu Asn Tyr Cys Ser His Asp
245 250 255

Ile His His Asp Asp Val Thr Arg Ala Val Gly Arg Thr Thr Ala Tyr
260 265 270

Asn Arg Val Ile Gly Asn Ser Tyr Thr Ala Ser Val Tyr Leu Gly Leu
275 280 285

Ala Ala Leu Leu Asp Gln Ala Asp Asp Leu Thr Gly Glu Arg Ile Gly

290

295

300

Phe Leu Ser Tyr Gly Ser Gly Ser Val Ala Glu Phe Phe Gly Gly Ile
 305 310 315 320

Val Val Ala Gly Tyr Arg Asp Arg Leu Arg Thr Ala Ala Asn Ile Glu
 325 330 335

Ala Val Ser Arg Arg Arg Pro Ile Asp Tyr Ala Gly Tyr Arg Glu Leu
 340 345 350

His Glu Trp Ala Phe Pro Ala Arg Arg Gly Ala His Ser Thr Pro Gln
 355 360 365

Gln Thr Thr Gly Pro Phe Arg Leu Ser Gly Ile Ser Gly His Lys Arg
 370 375 380

Leu Tyr Arg Ala Cys
 385

<210> 84

<211> 407

<212> PRT

<213> **Borrelia burgdorferi**

<400> 84

Met Arg Ile Gly Ile Ser Asp Ile Arg Ile Phe Leu Pro Leu Asn Tyr
 1 5 10 15

Leu Asp Phe Ser Val Leu Leu Glu Asn Pro Leu Tyr Phe Ser Asn Glu
 20 25 30

Val Phe Phe Lys Lys Ile Asn Arg Ala Ile Asp Ala Thr Leu Gln Lys
 35 40 45

Gly Phe Arg Phe Thr Ser Pro Asn Glu Asp Ser Val Thr Met Ala Ser
 50 55 60

Ser Ala Val Lys Leu Ile Phe Asp Asn Asn Leu Asp Leu Ser Lys
 65 70 75 80

Ile Arg Ile Leu Leu Gly Gly Thr Glu Thr Gly Val Asp His Ser Lys
 85 90 95

Ala Ile Ser Ser Tyr Val Phe Gly Ala Leu Lys Gln Ser Gly Ile Cys
 100 105 110

Leu Gly Asn Asn Phe Leu Thr Phe Gln Val Gln His Ala Cys Ala Gly
 115 120 125

Ala Ala Met Ser Leu His Thr Val Ala Ser Val Leu Ser His Ser Asn

130	135	140
Asn Ser Glu Tyr Gly Ile Val Phe Ser Ser Asp Ile Ala His Tyr Ser		
145	150	155
160		
Asn Leu Thr Thr Ala Glu Ile Thr Gln Gly Ala Gly Ala Thr Ala Ile		
165	170	175
Leu Ile Glu Lys Asn Pro Lys Ile Leu Ser Ile Asn Leu Ser Glu Phe		
180	185	190
Gly Val Tyr Thr Asp Asp Val Asp Asp Phe Phe Arg Pro Phe Gly Ser		
195	200	205
Val Glu Ala Lys Val Arg Gly Gln Tyr Ser Val Glu Cys Tyr Asn Asn		
210	215	220
Ala Asn Glu Asn Ala Leu Arg Asp Phe Ala Phe Lys Lys Gln Leu Ser		
225	230	235
240		
Met Lys Asp Leu Phe Ser Asn Tyr Arg Phe Val Leu His Val Pro Phe		
245	250	255
Ala Lys Met Pro Ile Asp Ser Met His Tyr Ile Leu Lys Lys Tyr Tyr		
260	265	270
Ser Asp Asp Glu Ser Val Arg Asn Ala Tyr Leu Glu Ser Ile Asp Phe		
275	280	285
Tyr Asp Gly Val Glu Ala Ala Met Glu Val Gly Asn Leu Tyr Thr Gly		
290	295	300
Ser Ile Phe Leu Ser Leu Ala Phe Tyr Leu Lys Arg Val Phe Ser Lys		
305	310	315
320		
Lys Asp Ile Thr Gly Glu Lys Ile Leu Phe Cys Ser Tyr Gly Ser Gly		
325	330	335
Asn Ile Met Ile Ile Tyr Glu Leu Thr Ile Glu Lys Ser Ala Phe Asp		
340	345	350
Val Ile Lys Leu Trp Asp Leu Glu Gly Leu Ile Lys Asn Arg Asn Asn		
355	360	365
Ala Asn Phe Glu Glu Tyr Lys Asp Phe Phe Gln Asn Lys Ile Ile Pro		
370	375	380
Gly Glu Ser Arg Gly Phe Tyr Leu Lys Glu Leu Arg Asn Asp Gly Tyr		
385	390	395
400		
Arg Val Tyr Gly Tyr Arg Ala		
405		

<210> 85

<211> 317

<212> PRT

<213> Streptococcus pneumoniae

<400> 85

Met Asp Arg Glu Pro Val Thr Val Arg Ser Tyr Ala Asn Ile Ala Ile
1 5 10 15

Ile Lys Tyr Trp Gly Lys Lys Glu Lys Glu Met Val Pro Ala Thr
20 25 30

Ser Ser Ile Ser Leu Thr Leu Glu Asn Met Tyr Thr Glu Thr Thr Leu
35 40 45

Ser Pro Leu Pro Ala Asn Val Thr Ala Asp Glu Phe Tyr Ile Asn Gly
50 55 60

Gln Leu Gln Asn Glu Val Glu His Ala Lys Met Ser Lys Ile Ile Asp
65 70 75 80

Arg Tyr Arg Pro Ala Gly Glu Gly Phe Val Arg Ile Asp Thr Gln Asn
85 90 95

Asn Met Pro Thr Ala Ala Gly Leu Ser Ser Ser Ser Gly Leu Ser
100 105 110

Ala Leu Val Lys Ala Cys Asn Ala Tyr Phe Lys Leu Gly Leu Asp Arg
115 120 125

Ser Gln Leu Ala Gln Glu Ala Lys Phe Ala Ser Gly Ser Ser Ser Arg
130 135 140

Ser Phe Tyr Gly Pro Leu Gly Ala Trp Asp Lys Asp Ser Gly Glu Ile
145 150 155 160

Tyr Pro Val Glu Thr Asp Leu Lys Leu Ala Met Ile Met Leu Val Leu
165 170 175

Glu Asp Lys Lys Lys Pro Ile Ser Ser Arg Asp Gly Met Lys Leu Cys
180 185 190

Val Glu Thr Ser Thr Phe Asp Asp Trp Val Arg Gln Ser Glu Lys
195 200 205

Asp Tyr Gln Asp Met Leu Ile Tyr Leu Lys Glu Asn Asp Phe Ala Lys
210 215 220

Ile Gly Glu Leu Thr Glu Lys Asn Ala Leu Ala Met His Ala Thr Thr
225 230 235 240

Lys Thr Ala Ser Pro Ala Phe Ser Tyr Leu Thr Asp Ala Ser Tyr Glu
245 250 255

Ala Met Ala Phe Val Arg Gln Leu Arg Glu Lys Gly Glu Ala Cys Tyr

260

265

270

Phe Thr Met Asp Ala Gly Pro Asn Val Lys Val Phe Cys Gln Glu Lys
 275 280 285

Asp Leu Glu His Leu Ser Glu Ile Phe Gly Gln Arg Tyr Arg Leu Ile
290 295 300

Val	Ser	Lys	Thr	Lys	Asp	Leu	Ser	Gln	Asp	Asp	Cys	Cys
305					310					315		

<210> 86

<211> 314

<212> PRT

<213> Streptococcus pyogenes

<400> 86

Met Asp Pro Asn Val Ile Thr Val Thr Ser Tyr Ala Asn Ile Ala Ile
1 5 10 15

Ile Lys Tyr Trp Gly Lys Glu Asn Gln Ala Lys Met Ile Pro Ser Thr
20 25 30

Ser Ser Ile Ser Leu Thr Leu Glu Asn Met Phe Thr Thr Thr Ser Val
35 40 45

Ser Phe Leu Pro Asp Thr Ala Thr Ser Asp Gln Phe Tyr Ile Asn Gly
50 55 60

Val Leu Gln Asn Asp Glu Glu His Thr Lys Ile Ser Thr Ile Ile Asp
65 70 75 80

Gln Phe Arg Gln Pro Gly Gln Ala Phe Val Lys Met Glu Thr Gln Asn
85 90 95

Asn	Met	Pro	Thr	Ala	Ala	Gly	Leu	Ser	Ser	Ser	Ser	Ser	Gly	Leu	Ser
				100				105						110	

Ala Leu Val Lys Ala Cys Asp Gln Leu Phe Asp Thr Gln Leu Asp Gln
115 120 125

Lys Ala Leu Ala Gln Lys Ala Lys Phe Ala Ser Gly Ser Ser Ser Arg
 130 135 140

Ser Phe Phe Gly Pro Val Ala Ala Trp Asp Lys Asp Ser Gly Ala Ile
 145 150 155 160

Tyr Lys Val Glu Thr Asp Leu Lys Met Ala Met Ile Met Leu Val Leu
 165 170 175

Asn Ala Ala Lys Lys Pro Ile Ser Ser Arg Glu Gly Met Lys Leu Cys

180

185

190

Arg Asp Thr Ser Thr Thr Phe Asp Glu Trp Val Glu Gln Ser Ala Ile
 195 200 205

Asp Tyr Gln His Met Leu Thr Tyr Leu Lys Thr Asn Asn Phe Glu Lys
 210 215 220

Val Gly Gln Leu Thr Glu Ala Asn Ala Leu Ala Met His Ala Thr Thr
 225 230 235 240

Lys Thr Ala Asn Pro Pro Phe Ser Tyr Leu Thr Lys Glu Ser Tyr Gln
 245 250 255

Ala Met Glu Ala Val Lys Glu Leu Arg Gln Glu Gly Phe Ala Cys Tyr
 260 265 270

Phe Thr Met Asp Ala Gly Pro Asn Val Lys Val Leu Cys Leu Glu Lys
 275 280 285

Asp Leu Ala Gln Leu Ala Glu Arg Leu Gly Lys Asn Tyr Arg Ile Ile
 290 295 300

Val Ser Lys Thr Lys Asp Leu Pro Asp Val
 305 310

<210> 87

<211> 331

<212> PRT

<213> Enterococcus faecalis

<400> 87

Met Leu Ser Gly Lys Ala Arg Ala His Thr Asn Ile Ala Leu Ile Lys
 1 5 10 15

Tyr Trp Gly Lys Ala Asn Glu Glu Tyr Ile Leu Pro Met Asn Ser Ser
 20 25 30

Leu Ser Leu Thr Leu Asp Ala Phe Tyr Thr Glu Thr Thr Val Thr Phe
 35 40 45

Asp Ala His Tyr Ser Glu Asp Val Phe Ile Leu Asn Gly Ile Leu Gln
 50 55 60

Asn Glu Lys Gln Thr Lys Lys Val Lys Glu Phe Leu Asn Leu Val Arg
 65 70 75 80

Gln Gln Ala Asp Cys Thr Trp Phe Ala Lys Val Glu Ser Gln Asn Phe
 85 90 95

Val Pro Thr Ala Ala Gly Leu Ala Ser Ser Ala Ser Gly Leu Ala Ala

100	105	110
Leu Ala Gly Ala Cys Asn Val Ala Leu Gly Leu Asn Leu Ser Ala Lys		
115	120	125
Asp Leu Ser Arg Leu Ala Arg Arg Gly Ser Gly Ser Ala Cys Arg Ser		
130	135	140
Ile Phe Gly Gly Phe Ala Gln Trp Asn Lys Gly His Ser Asp Glu Thr		
145	150	155
Ser Phe Ala Glu Asn Ile Pro Ala Asn Asn Trp Glu Asn Glu Leu Ala		
165	170	175
Met Leu Phe Ile Leu Ile Asn Asp Gly Glu Lys Asp Val Ser Ser Arg		
180	185	190
Asp Gly Met Lys Arg Thr Val Glu Thr Ser Ser Phe Tyr Gln Gly Trp		
195	200	205
Leu Asp Asn Val Glu Lys Asp Leu Ser Gln Val His Glu Ala Ile Lys		
210	215	220
Thr Lys Asp Phe Pro Arg Leu Gly Glu Ile Ile Glu Ala Asn Gly Leu		
225	230	235
Arg Met His Gly Thr Thr Leu Gly Ala Val Pro Pro Phe Thr Tyr Trp		
245	250	255
Ser Pro Gly Ser Leu Gln Ala Met Ala Leu Val Arg Gln Ala Arg Ala		
260	265	270
Lys Gly Ile Pro Cys Tyr Phe Thr Met Asp Ala Gly Pro Asn Val Lys		
275	280	285
Val Leu Val Glu Lys Lys Asn Leu Glu Ala Leu Lys Thr Phe Leu Ser		
290	295	300
Glu His Phe Ser Lys Glu Gln Leu Val Pro Ala Phe Ala Gly Pro Gly		
305	310	315
Ile Glu Leu Phe Glu Thr Lys Gly Met Asp Lys		
325	330	
<210> 88		
<211> 325		
<212> PRT		
<213> Enterococcus faecium		

<400> 88

Met Phe Lys Gly Lys Ala Arg Ala Tyr Thr Asn Ile Ala Leu Ile Lys

1

5

10

15

Tyr Trp Gly Lys Lys Asn Glu Glu Leu Ile Leu Pro Met Asn Asn Ser
 20 25 30

Leu Ser Leu Thr Leu Asp Ala Phe Tyr Thr Glu Thr Glu Val Ile Phe
 35 40 45

Ser Asp Ser Tyr Met Val Asp Glu Phe Tyr Leu Asp Gly Thr Leu Gln
 50 55 60

Asp Glu Lys Ala Thr Lys Lys Val Ser Gln Phe Leu Asp Leu Phe Arg
 65 70 75 80

Lys Glu Ala Gly Leu Ser Leu Lys Ala Ser Val Ile Ser Gln Asn Phe
 85 90 95

Val Pro Thr Ala Ala Gly Leu Ala Ser Ser Ala Ser Gly Leu Ala Ala
 100 105 110

Leu Ala Gly Ala Cys Asn Thr Ala Leu Lys Leu Gly Leu Asp Asp Leu
 115 120 125

Ser Leu Ser Arg Phe Ala Arg Arg Gly Ser Gly Ser Ala Cys Arg Ser
 130 135 140

Ile Phe Gly Gly Phe Val Glu Trp Glu Lys Gly His Asp Asp Leu Ser
 145 150 155 160

Ser Tyr Ala Lys Pro Val Pro Ser Asp Ser Phe Glu Asp Asp Leu Ala
 165 170 175

Met Val Phe Val Leu Ile Asn Asp Gln Lys Lys Glu Val Ser Ser Arg
 180 185 190

Asn Gly Met Arg Arg Thr Val Glu Thr Ser Asn Phe Tyr Gln Gly Trp
 195 200 205

Leu Asp Ser Val Glu Gly Asp Leu Tyr Gln Leu Lys Gln Ala Ile Lys
 210 215 220

Thr Lys Asp Phe Gln Leu Leu Gly Glu Thr Met Glu Arg Asn Gly Leu
 225 230 235 240

Lys Met His Gly Thr Thr Leu Ala Ala Gln Pro Pro Phe Thr Tyr Trp
 245 250 255

Ser Pro Asn Ser Leu Lys Ala Met Asp Ala Val Arg Gln Leu Arg Lys
 260 265 270

Gln Gly Ile Pro Cys Tyr Phe Thr Met Asp Ala Gly Pro Asn Val Lys
 275 280 285

Val Leu Val Glu Asn Ser His Leu Ser Glu Val Gln Glu Thr Phe Thr
 290 295 300

Lys Leu Phe Ser Lys Glu Gln Val Ile Thr Ala His Ala Gly Pro Gly

305

310

315

320

Ile Ala Ile Ile Glu
325

<210> 89

<211> 327

<212> PRT

<213> *Staphylococcus haemolyticus*

<400> 89

Met Lys Lys Ser Gly Lys Ala Arg Ala His Thr Asn Ile Ala Leu Ile
1 5 10 15

Lys Tyr Trp Gly Lys Ala Asp Glu Ala Leu Ile Ile Pro Met Asn Asn
20 25 30

Ser Leu Ser Val Thr Leu Asp Arg Phe Tyr Thr Glu Thr Arg Val Thr
35 40 45

Phe Asp Glu Thr Leu Thr Glu Asp Gln Leu Ile Leu Asn Gly Glu Ala
50 55 60

Val Asn Ala Lys Glu Ser Ala Lys Ile Gln Arg Tyr Met Glu Met Ile
65 70 75 80

Arg Lys Glu Ala Gly Ile Ser His Glu Ala Leu Ile Glu Ser Glu Asn
85 90 95

Phe Val Pro Thr Ala Ala Gly Leu Ala Ser Ser Ala Ser Ala Tyr Ala
100 105 110

Ala Leu Ala Gly Ala Cys Asn Glu Ala Leu Gln Leu Gly Leu Ser Asp
115 120 125

Lys Asp Leu Ser Arg Leu Ala Arg Arg Gly Ser Gly Ser Ala Ser Arg
130 135 140

Ser Ile Tyr Gly Gly Phe Ala Glu Trp Glu Lys Gly Asn Asp Asp Glu
145 150 155 160

Thr Ser Phe Ala His Arg Val Glu Ala Asp Gly Trp Glu Asn Glu Leu
165 170 175

Ala Met Val Phe Val Val Ile Asn Asn Lys Ser Lys Lys Val Ser Ser
180 185 190

Arg Ser Gly Met Ser Leu Thr Arg Asp Thr Ser Arg Phe Tyr Gln Tyr
195 200 205

Trp Leu Asp Asn Val Glu Pro Asp Leu Lys Glu Thr Lys Glu Ala Ile

210

215

220

Ala Gln Lys Asp Phe Lys Arg Met Gly Glu Val Ile Glu Ala Asn Gly
 225 230 235 240

Leu Arg Met His Ala Thr Asn Leu Gly Ala Gln Pro Pro Phe Thr Tyr
 245 250 255

Leu Val Pro Glu Ser Tyr Asp Ala Met Arg Ile Val His Glu Cys Arg
 260 265 270

Glu Ala Gly Leu Pro Cys Tyr Phe Thr Met Asp Ala Gly Pro Asn Val
 275 280 285

Lys Val Leu Ile Glu Lys Lys Asn Gln Gln Ala Ile Val Asp Lys Phe
 290 295 300

Leu Gln Glu Phe Asp Gln Ser Gln Ile Ile Thr Ser Asp Ile Thr Gln
 305 310 315 320

Ser Gly Val Glu Ile Ile Lys
 325

<210> 90

<211> 327

<212> PRT

<213> *Staphylococcus epidermidis*

<400> 90

Met Val Lys Ser Gly Lys Ala Arg Ala His Thr Asn Ile Ala Leu Ile
 1 5 10 15

Lys Tyr Trp Gly Lys Ala Asp Glu Thr Tyr Ile Ile Pro Met Asn Asn
 20 25 30

Ser Leu Ser Val Thr Leu Asp Arg Phe Tyr Thr Glu Thr Lys Val Thr
 35 40 45

Phe Asp Pro Asp Phe Thr Glu Asp Cys Leu Ile Leu Asn Gly Asn Glu
 50 55 60

Val Asn Ala Lys Glu Lys Glu Lys Ile Gln Asn Tyr Met Asn Ile Val
 65 70 75 80

Arg Asp Leu Ala Gly Asn Arg Leu His Ala Arg Ile Glu Ser Glu Asn
 85 90 95

Tyr Val Pro Thr Ala Ala Gly Leu Ala Ser Ser Ala Ser Ala Tyr Ala
 100 105 110

Ala Leu Ala Ala Ala Cys Asn Glu Ala Leu Ser Leu Asn Leu Ser Asp

115	120	125
Thr Asp Leu Ser Arg Leu Ala Arg Arg Gly Ser Gly Ser Ala Ser Arg		
130	135	140
Ser Ile Phe Gly Gly Phe Ala Glu Trp Glu Lys Gly His Asp Asp Leu		
145	150	155
160		
Thr Ser Tyr Ala His Gly Ile Asn Ser Asn Gly Trp Glu Lys Asp Leu		
165	170	175
Ser Met Ile Phe Val Val Ile Asn Asn Gln Ser Lys Lys Val Ser Ser		
180	185	190
Arg Ser Gly Met Ser Leu Thr Arg Asp Thr Ser Arg Phe Tyr Gln Tyr		
195	200	205
Trp Leu Asp His Val Asp Glu Asp Leu Asn Glu Ala Lys Glu Ala Val		
210	215	220
Lys Asn Gln Asp Phe Gln Arg Leu Gly Glu Val Ile Glu Ala Asn Gly		
225	230	235
240		
Leu Arg Met His Ala Thr Asn Leu Gly Ala Gln Pro Pro Phe Thr Tyr		
245	250	255
Leu Val Gln Glu Ser Tyr Asp Ala Met Ala Ile Val Glu Gln Cys Arg		
260	265	270
Lys Ala Asn Leu Pro Cys Tyr Phe Thr Met Asp Ala Gly Pro Asn Val		
275	280	285
Lys Val Leu Val Glu Lys Lys Asn Lys Gln Ala Val Met Glu Gln Phe		
290	295	300
Leu Lys Val Phe Asp Glu Ser Lys Ile Ile Ala Ser Asp Ile Ile Ser		
305	310	315
320		
Ser Gly Val Glu Ile Ile Lys		
325		
<210> 91		
<211> 327		
<212> PRT		
<213> Staphylococcus aureus		
<400> 91		
Met Ile Lys Ser Gly Lys Ala Arg Ala His Thr Asn Ile Ala Leu Ile		
1	5	10
15		
Lys Tyr Trp Gly Lys Lys Asp Glu Ala Leu Ile Ile Pro Met Asn Asn		

20

25

30

Ser Ile Ser Val Thr Leu Glu Lys Phe Tyr Thr Glu Thr Lys Val Thr
 35 40 45

Phe Asn Asp Gln Leu Thr Gln Asp Gln Phe Trp Leu Asn Gly Glu Lys
 50 55 60

Val Ser Gly Lys Glu Leu Glu Lys Ile Ser Lys Tyr Met Asp Ile Val
 65 70 75 80

Arg Asn Arg Ala Gly Ile Asp Trp Tyr Ala Glu Ile Glu Ser Asp Asn
 85 90 95

Phe Val Pro Thr Ala Ala Gly Leu Ala Ser Ser Ala Ser Ala Tyr Ala
 100 105 110

Ala Leu Ala Ala Ala Cys Asn Gln Ala Leu Asp Leu Gln Leu Ser Asp
 115 120 125

Lys Asp Leu Ser Arg Leu Ala Arg Ile Gly Ser Gly Ser Ala Ser Arg
 130 135 140

Ser Ile Tyr Gly Gly Phe Ala Glu Trp Glu Lys Gly Tyr Asn Asp Glu
 145 150 155 160

Thr Ser Tyr Ala Val Pro Leu Glu Ser Asn His Phe Glu Asp Asp Leu
 165 170 175

Ala Met Ile Phe Val Val Ile Asn Gln His Ser Lys Lys Val Pro Ser
 180 185 190

Arg Tyr Gly Met Ser Leu Thr Arg Asn Thr Ser Arg Phe Tyr Gln Tyr
 195 200 205

Trp Leu Asp His Ile Asp Glu Asp Leu Ala Glu Ala Lys Ala Ala Ile
 210 215 220

Gln Asp Lys Asp Phe Lys Arg Leu Gly Glu Val Ile Glu Glu Asn Gly
 225 230 235 240

Leu Arg Met His Ala Thr Asn Leu Gly Ser Thr Pro Pro Phe Thr Tyr
 245 250 255

Leu Val Gln Glu Ser Tyr Asp Val Met Ala Leu Val His Glu Cys Arg
 260 265 270

Glu Ala Gly Tyr Pro Cys Tyr Phe Thr Met Asp Ala Gly Pro Asn Val
 275 280 285

Lys Ile Leu Val Glu Lys Lys Asn Lys Gln Gln Ile Ile Asp Lys Leu
 290 295 300

Leu Thr Gln Phe Asp Asn Asn Gln Ile Ile Asp Ser Asp Ile Ile Ala
 305 310 315 320

Thr Gly Ile Glu Ile Ile Glu

325

<210> 92

<211> 350

<212> PRT

<213> Streptomyces sp. CL190

<400> 92

Met Arg Ser Glu His Pro Thr Thr Thr Val Leu Gln Ser Arg Glu Gln
1 5 10 15

Gly Ser Ala Ala Gly Ala Thr Ala Val Ala His Pro Asn Ile Ala Leu
20 25 30

Ile Lys Tyr Trp Gly Lys Arg Asp Glu Arg Leu Ile Leu Pro Cys Thr
35 40 45

Thr Ser Leu Ser Met Thr Leu Asp Val Phe Pro Thr Thr Thr Glu Val
50 55 60

Arg Leu Asp Pro Ala Ala Glu His Asp Thr Ala Ala Leu Asn Gly Glu
65 70 75 80

Val Ala Thr Gly Glu Thr Leu Arg Arg Ile Ser Ala Phe Leu Ser Leu
85 90 95

Val Arg Glu Val Ala Gly Ser Asp Gln Arg Ala Val Val Asp Thr Arg
100 105 110

Asn Thr Val Pro Thr Gly Ala Gly Leu Ala Ser Ser Ala Ser Gly Phe
115 120 125

Ala Ala Leu Ala Val Ala Ala Ala Ala Tyr Gly Leu Glu Leu Asp
130 135 140

Asp Arg Gly Leu Ser Arg Leu Ala Arg Arg Gly Ser Gly Ser Ala Ser
145 150 155 160

Arg Ser Ile Phe Gly Phe Ala Val Trp His Ala Gly Pro Asp Gly
165 170 175

Thr Ala Thr Glu Ala Asp Leu Gly Ser Tyr Ala Glu Pro Val Pro Ala
180 185 190

Ala Asp Leu Asp Pro Ala Leu Val Ile Ala Val Val Asn Ala Gly Pro
195 200 205

Lys Pro Val Ser Ser Arg Glu Ala Met Arg Arg Thr Val Asp Thr Ser
210 215 220

Pro Leu Tyr Arg Pro Trp Ala Asp Ser Ser Lys Asp Asp Leu Asp Glu

225	230	235	240
Met Arg Ser Ala Leu Leu Arg Gly Asp	Leu Glu Ala Val Gly Glu Ile		
245	250		255
Ala Glu Arg Asn Ala Leu Gly Met His Ala Thr Met Leu Ala Ala Arg			
260	265		270
Pro Ala Val Arg Tyr Leu Ser Pro Ala Thr Val Thr Val Leu Asp Ser			
275	280		285
Val Leu Gln Ieu Arg Lys Asp Gly Val Leu Ala Tyr Ala Thr Met Asp			
290	295		300
Ala Gly Pro Asn Val Lys Val Leu Cys Arg Arg Ala Asp Ala Glu Arg			
305	310		320
Val Ala Asp Val Val Arg Ala Ala Ala Ser Gly Gly Gln Val Leu Val			
325	330		335
Ala Gly Pro Gly Asp Gly Ala Arg Leu Leu Ser Glu Gly Ala			
340	345		350
<210> 93			
<211> 331			
<212> PRT			
<213> Streptomyces griseolosporeus			
<400> 93			
Ala Thr Ala Val Ala Gln Pro Asn Ile Ala Leu Ile Lys Tyr Trp Gly			
1	5	10	15
Lys Lys Asp Glu His Leu Val Leu Pro Arg Thr Asp Ser Leu Ser Met			
20	25		30
Thr Leu Asp Ile Phe Pro Thr Thr Arg Val Gln Leu Ala Pro Gly			
35	40		45
Ala Gly Gln Asp Thr Val Ala Phe Asn Gly Glu Pro Ala Thr Gly Glu			
50	55		60
Ala Glu Arg Arg Ile Thr Ala Phe Leu Arg Leu Val Arg Glu Arg Ser			
65	70		80
Gly Arg Thr Glu Arg Ala Arg Val Glu Thr Glu Asn Thr Val Pro Thr			
85	90		95
Gly Ala Gly Leu Ala Ser Ser Ala Ser Gly Phe Ala Ala Leu Ala Val			
100	105		110
Ala Ala Ala Ala Tyr Gly Leu Gly Leu Asp Ala Arg Gly Leu Ser			

115	120	125
Arg Leu Ala Arg Arg Gly Ser Gly Ser Ala Ser Arg Ser Ile Phe Asp		
130	135	140
Gly Phe Ala Val Trp His Ala Gly His Ala Gly Gly Thr Pro Glu Glu		
145	150	155
160		
Ala Asp Leu Gly Ser Tyr Ala Glu Pro Val Pro Ala Val Asp Leu Glu		
165	170	175
Pro Ala Leu Val Val Ala Val Val Ser Ala Ala Pro Lys Ala Val Ser		
180	185	190
Ser Arg Glu Ala Met Arg Arg Thr Val Asp Thr Ser Pro Leu Tyr Glu		
195	200	205
Pro Trp Ala Val Ser Ser Arg Ala Asp Leu Ala Asp Ile Gly Ala Ala		
210	215	220
Leu Ala Arg Gly Asn Leu Pro Ala Val Gly Glu Ile Ala Glu Arg Asn		
225	230	235
240		
Ala Leu Gly Met His Ala Thr Met Leu Ala Ala Arg Pro Ala Val Arg		
245	250	255
Tyr Leu Ser Pro Ala Ser Leu Ala Val Leu Asp Gly Val Leu Gln Leu		
260	265	270
Arg Arg Asp Gly Val Pro Ala Tyr Ala Thr Met Asp Ala Gly Pro Asn		
275	280	285
Val Lys Val Leu Cys Pro Arg Ser Asp Ala Glu Arg Val Ala Glu Ala		
290	295	300
Leu Arg Ala Ala Ala Pro Val Gly Ala Val His Ile Ala Gly Pro Gly		
305	310	315
320		
Arg Gly Ala Arg Leu Val Ala Glu Glu Cys Arg		
325	330	
<210> 94		
<211> 312		
<212> PRT		
<213> Borrelia burgdorferi		

<400> 94

Met Lys Ile Lys Cys Lys Val His Ala Ser Leu Ala Leu Ile Lys Tyr
1 5 10 15

Trp Gly Lys Lys Asp Val Phe Leu Asn Ile Pro Ala Thr Ser Ser Leu

20

25

30

Ala Val Ser Val Asp Lys Phe Tyr Ser Ile Ser Glu Leu Glu Leu Ser
 35 40 45

Asn Arg Asp Glu Ile Ile Leu Asn Ser Lys Pro Val Ile Leu Lys Asn
 50 55 60

Arg Glu Lys Val Phe Phe Asp Tyr Ala Arg Lys Ile Leu Asn Glu Pro
 65 70 75 80

Asn Val Arg Phe Lys Ile Lys Ser Lys Asn Asn Phe Pro Thr Ala Ala
 85 90 95

Gly Leu Ala Ser Ser Ser Gly Phe Ala Ser Ile Ala Ala Cys Ile
 100 105 110

Leu Lys Tyr Phe Asn Lys Tyr Ser Cys Asn Ser Ala Ser Asn Leu Ala
 115 120 125

Arg Val Gly Ser Ala Ser Ala Ala Arg Ala Ile Tyr Gly Gly Phe Thr
 130 135 140

Ile Leu Lys Glu Gly Ser Lys Glu Ser Phe Gln Leu Arg Asp Gln Ser
 145 150 155 160

Tyr Phe Asn Asp Leu Arg Ile Ile Phe Ala Ile Ile Asp Ser Asn Glu
 165 170 175

Lys Glu Leu Ser Ser Arg Ala Ala Met Asn Ile Cys Lys Arg His Lys
 180 185 190

Phe Tyr Tyr Asp Ala Trp Ile Ala Ser Ser Lys Lys Ile Phe Lys Asp
 195 200 205

Ala Leu Tyr Phe Phe Leu Lys Lys Asp Phe Ile His Phe Gly Ala Thr
 210 215 220

Ile Val Lys Ser Tyr Gln Asn Met Phe Ala Leu Met Phe Ala Ser Ser
 225 230 235 240

Ile Phe Tyr Phe Lys Asn Ser Thr Ile Asp Leu Ile Arg Tyr Ala Ala
 245 250 255

Asp Leu Arg Asn Glu Gly Ile Phe Val Phe Glu Thr Met Asp Ala Gly
 260 265 270

Pro Gln Val Lys Phe Leu Cys Leu Glu Glu Asn Leu Asn Thr Ile Leu
 275 280 285

Lys Gly Leu Lys Gln Asn Phe Thr Gly Ile Asp Phe Ile Val Ser Lys
 290 295 300

Val Gly Cys Asp Leu Glu Trp Ile
 305 310

<210> 95

<211> 292

<212> PRT

<213> Streptococcus pneumoniae

<400> 95

Met Thr Lys Lys Val Gly Val Gly Gln Ala His Ser Lys Ile Ile Leu
1 5 10 15

Ile Gly Glu His Ala Val Val Tyr Gly Tyr Pro Ala Ile Ser Leu Pro
20 25 30

Leu Leu Glu Val Glu Val Thr Cys Lys Val Val Ser Ala Glu Ser Pro
35 40 45

Trp Arg Leu Tyr Glu Glu Asp Thr Leu Ser Met Ala Val Tyr Ala Ser
50 55 60

Leu Glu Tyr Leu Asp Ile Thr Glu Ala Cys Val Arg Cys Glu Ile Asp
65 70 75 80

Ser Ala Ile Pro Glu Lys Arg Gly Met Gly Ser Ser Ala Ala Ile Ser
85 90 95

Ile Ala Ala Ile Arg Ala Val Phe Asp Tyr Tyr Gln Ala Asp Leu Pro
100 105 110

His Asp Val Leu Glu Ile Leu Val Asn Arg Ala Glu Met Ile Ala His
115 120 125

Met Asn Pro Ser Gly Leu Asp Ala Lys Thr Cys Leu Ser Asp Gln Pro
130 135 140

Ile Arg Phe Ile Lys Asn Val Gly Phe Thr Glu Leu Glu Met Asp Leu
145 150 155 160

Ser Ala Tyr Leu Val Ile Ala Asp Thr Gly Val Tyr Gly His Thr Arg
165 170 175

Glu Ala Ile Gln Val Val Gln Asn Lys Gly Lys Asp Ala Leu Pro Phe
180 185 190

Leu His Ala Leu Gly Glu Leu Thr Gln Gln Ala Glu Val Ala Ile Ser
195 200 205

Gln Lys Tyr Ala Glu Gly Leu Gly Leu Ile Phe Ser Gln Ala His Leu
210 215 220

His Leu Lys Glu Ile Gly Val Ser Ser Pro Glu Ala Asp Phe Leu Val
225 230 235 240

Glu Thr Ala Leu Ser Tyr Gly Ala Leu Gly Ala Lys Met Ser Gly Gly

245

250

255

Gly Leu Gly Gly Cys Ile Ile Ala Leu Val Thr Asn Leu Thr His Ala
260 265 270

Gln Glu Leu Ala Glu Arg Leu Glu Glu Lys Gly Ala Val Gln Thr Trp
275 280 285

Ile Glu Ser Leu
290

<210> 96

<211> 292

<212> PRT

<213> Streptococcus pyogenes

<400> 96

Met Asn Glu Asn Ile Gly Tyr Gly Lys Ala His Ser Lys Ile Ile Leu
1 5 10 15

Ile Gly Glu His Ala Val Val Tyr Gly Tyr Pro Ala Ile Ala Leu Pro
20 25 30

Leu Thr Asp Ile Glu Val Val Cys His Ile Phe Pro Ala Asp Lys Pro
35 40 45

Leu Val Phe Asp Phe Tyr Asp Thr Leu Ser Thr Ala Ile Tyr Ala Ala
50 55 60

Leu Asp Tyr Leu Gln Arg Leu Gln Glu Pro Ile Ala Tyr Glu Ile Val
65 70 75 80

Ser Gln Val Pro Gln Lys Arg Gly Met Gly Ser Ser Ala Ala Val Ser
85 90 95

Ile Ala Ala Ile Arg Ala Val Phe Ser Tyr Cys Gln Glu Pro Leu Ser
100 105 110

Asp Asp Leu Leu Glu Ile Leu Val Asn Lys Ala Glu Ile Ile Ala His
115 120 125

Thr Asn Pro Ser Gly Leu Asp Ala Lys Thr Cys Leu Ser Asp His Ala
130 135 140

Ile Lys Phe Ile Arg Asn Ile Gly Phe Glu Thr Ile Glu Ile Ala Leu
145 150 155 160

Asn Gly Tyr Leu Ile Ile Ala Asp Thr Gly Ile His Gly His Thr Arg
165 170 175

Glu Ala Val Asn Lys Val Ala Gln Phe Glu Glu Thr Asn Leu Pro Tyr

180

185

190

Leu Ala Lys Leu Gly Ala Leu Thr Gln Ala Leu Glu Arg Ala Ile Asn
 195 200 205

Gln Lys Asn Lys Val Ala Ile Gly Gln Leu Met Thr Gln Ala His Ser
 210 215 220

Ala Leu Lys Ala Ile Gly Val Ser Ile Ser Lys Ala Asp Gln Leu Val
 225 230 235 240

Glu Ala Ala Leu Arg Ala Gly Ala Leu Gly Ala Lys Met Thr Gly Gly
 245 250 255

Gly Leu Gly Gly Cys Met Ile Ala Leu Ala Asp Thr Lys Asp Met Ala
 260 265 270

Glu Lys Ile Ser His Lys Leu Lys Glu Glu Gly Ala Val Asn Thr Trp
 275 280 285

Ile Gln Met Leu
 290

<210> 97

<211> 314

<212> PRT

<213> Enterococcus faecalis

<400> 97

Met Asn Ile Lys Lys Gln Gly Leu Gly Gln Ala Thr Gly Lys Ile Ile
 1 5 10 15

Leu Met Gly Glu His Ala Val Val Tyr Gly Glu Pro Ala Ile Ala Phe
 20 25 30

Pro Phe Gln Ala Thr Glu Ile Thr Ala Val Phe Thr Pro Ala Lys Thr
 35 40 45

Met Gln Ile Asp Cys Ala Tyr Phe Thr Gly Leu Leu Glu Asp Val Pro
 50 55 60

Gln Glu Leu Ala Asn Ile Lys Glu Val Val Gln Gln Thr Leu His Phe
 65 70 75 80

Leu Lys Glu Asp Thr Phe Lys Gly Thr Leu Thr Leu Thr Ser Thr Ile
 85 90 95

Pro Ala Glu Arg Gly Met Gly Ser Ser Ala Ala Thr Ala Val Ala Ile
 100 105 110

Val Arg Ser Leu Phe Asp Tyr Phe Asp Tyr Ala Tyr Thr Tyr Gln Glu

115

120

125

Leu Phe Glu Leu Val Ser Leu Ser Glu Lys Ile Ala His Gly Asn Pro
 130 135 140

Ser Gly Ile Asp Ala Ala Ala Thr Ser Gly Ala Asp Pro Leu Phe Phe
 145 150 155 160

Thr Arg Gly Phe Pro Pro Thr His Phe Ser Met Asn Leu Ser Asn Ala
 165 170 175

Tyr Leu Val Val Ala Asp Thr Gly Ile Lys Gly Gln Thr Arg Glu Ala
 180 185 190

Val Lys Asp Ile Ala Gln Leu Ala Gln Asn Asn Pro Thr Ala Ile Ala
 195 200 205

Glu Thr Met Lys Gln Leu Gly Ser Phe Thr Lys Glu Ala Lys Gln Ala
 210 215 220

Ile Leu Gln Asp Asp Lys Gln Lys Leu Gly Gln Leu Met Thr Leu Ala
 225 230 235 240

Gln Glu Gln Leu Gln Gln Leu Thr Val Ser Asn Asp Met Leu Asp Arg
 245 250 255

Leu Val Ala Leu Ser Leu Glu His Gly Ala Leu Gly Ala Lys Leu Thr
 260 265 270

Gly Gly Gly Arg Gly Gly Cys Met Ile Ala Leu Thr Asp Asn Lys Lys
 275 280 285

Thr Ala Gln Thr Ile Ala Gln Thr Leu Glu Glu Asn Gly Ala Val Ala
 290 295 300

Thr Trp Ile Gln Ser Leu Glu Val Lys Lys
 305 310

<210> 98

<211> 314

<212> PRT

<213> Enterococcus faecium

<400> 98

Met Ala Asn Tyr Gly Gln Gly Glu Ser Ser Gly Lys Ile Ile Leu Met
 1 5 10 15

Gly Glu His Ala Val Val Tyr Gly Glu Pro Ala Ile Ala Phe Pro Phe
 20 25 30

Tyr Ala Thr Lys Val Thr Ala Phe Leu Glu Leu Asp Ala Met Asp

35

40

45

Asp Gln Leu Val Ser Ser Tyr Tyr Ser Gly Asn Leu Ala Glu Ala Pro
 50 55 60

His Ala Leu Lys Asn Ile Lys Lys Leu Phe Ile His Leu Lys Lys Gln
 65 70 75 80

His Asp Ile Gln Lys Asn Leu Gln Leu Thr Ile Glu Ser Thr Ile Pro
 85 90 95

Ala Glu Arg Gly Met Gly Ser Ser Ala Ala Val Ala Thr Ala Val Thr
 100 105 110

Arg Ala Phe Tyr Asp Tyr Leu Ala Phe Pro Leu Ser Arg Glu Ile Leu
 115 120 125

Leu Glu Asn Val Gln Leu Ser Glu Lys Ile Ala His Gly Asn Pro Ser
 130 135 140

Gly Ile Asp Ala Ala Ala Thr Ser Ser Leu Gln Pro Ile Tyr Phe Thr
 145 150 155 160

Lys Gly His Pro Phe Asp Tyr Phe Ser Leu Asn Ile Asp Ala Phe Leu
 165 170 175

Ile Val Ala Asp Thr Gly Ile Lys Gly Gln Thr Arg Glu Ala Val Lys
 180 185 190

Asp Val Ala His Leu Phe Glu Thr Gln Pro His Glu Thr Gly Gln Met
 195 200 205

Ile Gln Lys Leu Gly Tyr Leu Thr Lys Gln Ala Lys Gln Ala Ile Ile
 210 215 220

Glu Asn Ser Pro Glu Thr Leu Ala Gln Thr Met Asp Glu Ser Gln Ser
 225 230 235 240

Leu Leu Glu Lys Leu Thr Ile Ser Asn Asp Phe Leu Asn Leu Leu Ile
 245 250 255

Gln Thr Ala Lys Asp Thr Gly Ala Leu Gly Ala Lys Leu Thr Gly Gly
 260 265 270

Gly Arg Gly Gly Cys Met Ile Ala Leu Ala Gln Thr Lys Thr Lys Ala
 275 280 285

Gln Glu Ile Ser Gln Ala Leu Glu Asp Ala Gly Ala Ala Glu Thr Trp
 290 295 300

Ile Gln Gly Leu Gly Val His Thr Tyr Val
 305 310

<210> 99

<211> 307

<212> PRT

<213> *Staphylococcus haemolyticus*

<400> 99

Met Val Gln Arg Gly Tyr Gly Glu Ser Asn Gly Lys Ile Ile Leu Ile
1 5 10 15

Gly Glu His Ala Val Thr Phe Gly Glu Pro Ala Ile Ala Ile Pro Phe
20 25 30

Thr Ser Gly Lys Val Lys Val Leu Ile Glu Ser Leu Glu Lys Gly Asn
35 40 45

Tyr Ser Ala Ile Gln Ser Asp Val Tyr Asp Gly Pro Leu Tyr Asp Ala
50 55 60

Pro Glu His Leu Lys Ser Leu Ile Gly His Phe Val Glu Asn Lys Lys
65 70 75 80

Val Glu Glu Pro Leu Leu Ile Lys Ile Gln Ala Asn Leu Pro Pro Ser
85 90 95

Arg Gly Leu Gly Ser Ser Ala Ala Val Ala Val Ala Phe Ile Arg Ala
100 105 110

Ser Tyr Asp Tyr Leu Gly Leu Pro Leu Thr Asp Lys Glu Leu Leu Glu
115 120 125

Asn Ala Asp Trp Ala Glu Arg Ile Ala His Gly Lys Pro Ser Gly Ile
130 135 140

Asp Thr Lys Thr Ile Val Thr Asn Gln Pro Val Trp Tyr Gln Lys Gly
145 150 155 160

Glu Val Glu Ile Leu Lys Thr Leu Asp Leu Asp Gly Tyr Met Val Val
165 170 175

Ile Asp Thr Gly Val Lys Gly Ser Thr Lys Gln Ala Val Glu Asp Val
180 185 190

His Gln Leu Cys Asp Asn Asp Lys Asn Tyr Met Gln Val Val Lys His
195 200 205

Ile Gly Ser Leu Val Tyr Ser Ala Ser Glu Ala Ile Glu His His Ser
210 215 220

Phe Asp Gln Leu Ala Thr Ile Phe Asn Gln Cys Gln Asp Asp Leu Arg
225 230 235 240

Thr Leu Thr Val Ser His Asp Lys Ile Glu Met Phe Leu Arg Leu Gly
245 250 255

Glu Glu Asn Gly Ser Val Ala Gly Lys Leu Thr Gly Gly Arg Gly

260	265	270	
Gly Ser Met Leu Ile Leu Ala Lys Glu Leu Gln Thr Ala Lys Asn Ile			
275	280	285	
Val Ala Ala Val Glu Lys Ala Gly Ala Gln His Thr Trp Ile Glu Lys			
290	295	300	
Leu Gly Gly			
305			
<210> 100			
<211> 306			
<212> PRT			
<213> Staphylococcus epidermidis			
<400> 100			
Met Thr Arg Gln Gly Tyr Gly Glu Ser Thr Gly Lys Ile Ile Leu Met			
1	5	10	15
Gly Glu His Ala Val Thr Phe Gly Gln Pro Ala Ile Ala Ile Pro Phe			
20	25	30	
Asn Ala Gly Lys Ile Lys Val Leu Ile Glu Ser Leu Asp Glu Gly Asn			
35	40	45	
Tyr Ser Ser Ile Thr Ser Asp Val Tyr Asp Gly Met Leu Tyr Asp Ala			
50	55	60	
Pro Glu His Leu Lys Ser Ile Ile Asn Arg Phe Val Glu Lys Ser Gly			
65	70	75	80
Val Lys Glu Pro Leu Ser Val Lys Ile Gln Thr Asn Leu Pro Pro Ser			
85	90	95	
Arg Gly Leu Gly Ser Ser Ala Ala Val Ala Val Ala Phe Val Arg Ala			
100	105	110	
Ser Tyr Asp Phe Met Asp Gln Pro Leu Asp Asp Lys Thr Leu Ile Lys			
115	120	125	
Glu Ala Asn Trp Ala Glu Gln Ile Ala His Gly Lys Pro Ser Gly Ile			
130	135	140	
Asp Thr Gln Thr Ile Val Ser Asn Lys Pro Val Trp Phe Lys Gln Gly			
145	150	155	160
Gln Ala Glu Thr Leu Lys Ser Leu Lys Leu Asn Gly Tyr Met Val Val			
165	170	175	
Ile Asp Thr Gly Val Lys Gly Ser Thr Lys Gln Ala Val Glu Asp Val			

180

185

190

His Val Leu Cys Glu Ser Asp Glu Tyr Met Lys Tyr Ile Glu His Ile
 195 200 205

Gly Thr Leu Val His Ser Ala Ser Glu Ser Ile Glu Gln His Asp Phe
 210 215 220

His His Leu Ala Asp Ile Phe Asn Ala Cys Gln Glu Asp Leu Arg His
 225 230 235 240

Leu Thr Val Ser His Asp Lys Ile Glu Lys Leu Leu Gln Ile Gly Lys
 245 250 255

Glu His Gly Ala Ile Ala Gly Lys Leu Thr Gly Gly Gly Arg Gly Gly
 260 265 270

Ser Met Leu Leu Ala Glu Asn Leu Lys Thr Ala Lys Thr Ile Val
 275 280 285

Ala Ala Val Glu Lys Ala Gly Ala Ala His Thr Trp Ile Glu His Leu
 290 295 300

Gly Gly
 305

<210> 101

<211> 306

<212> PRT

<213> Staphylococcus aureus

<400> 101

Met Thr Arg Lys Gly Tyr Gly Glu Ser Thr Gly Lys Ile Ile Leu Ile
 1 5 10 15

Gly Glu His Ala Val Thr Phe Gly Glu Pro Ala Ile Ala Val Pro Phe
 20 25 30

Asn Ala Gly Lys Ile Lys Val Leu Ile Glu Ala Leu Glu Ser Gly Asn
 35 40 45

Tyr Ser Ser Ile Lys Ser Asp Val Tyr Asp Gly Met Leu Tyr Asp Ala
 50 55 60

Pro Asp His Leu Lys Ser Leu Val Asn Arg Phe Val Glu Leu Asn Asn
 65 70 75 80

Ile Thr Glu Pro Leu Ala Val Thr Ile Gln Thr Asn Leu Pro Pro Ser
 85 90 95

Arg Gly Leu Gly Ser Ser Ala Ala Val Ala Val Ala Phe Val Arg Ala

100 105 110

Ser Tyr Asp Phe Leu Gly Lys Ser Leu Thr Lys Glu Glu Leu Ile Glu
115 120 125

Lys Ala Asn Trp Ala Glu Gln Ile Ala His Gly Lys Pro Ser Gly Ile
130 135 140

Asp Thr Gln Thr Ile Val Ser Gly Lys Pro Val Trp Phe Gln Lys Gly
145 150 155 160

Gln Ala Glu Thr Leu Lys Thr Leu Ser Leu Asp Gly Tyr Met Val Val
165 170 175

Ile Asp Thr Gly Val Lys Gly Ser Thr Arg Gln Ala Val Glu Asp Val
180 185 190

His Lys Leu Cys Glu Asp Pro Gln Tyr Met Ser His Val Lys His Ile
195 200 205

Gly Lys Leu Val Leu Arg Ala Ser Asp Val Ile Glu His His Asn Phe
210 215 220

Glu Ala Leu Ala Asp Ile Phe Asn Glu Cys His Ala Asp Leu Lys Ala
225 230 235 240

Leu Thr Val Ser His Asp Lys Ile Glu Gln Leu Met Lys Ile Gly Lys
245 250 255

Glu Asn Gly Ala Ile Ala Gly Lys Leu Thr Gly Ala Gly Arg Gly Gly
260 265 270

Ser Met Leu Leu Ala Lys Asp Leu Pro Thr Ala Lys Asn Ile Val
275 280 285

Lys Ala Val Glu Lys Ala Gly Ala Ala His Thr Trp Ile Glu Asn Leu
290 295 300

Gly Gly
305

<210> 102

<211> 345

<212> PRT

<213> Streptomyces sp. CL190

<400> 102

Met Gln Lys Arg Gln Arg Glu Leu Ser Ala Leu Thr Leu Pro Thr Ser
1 5 10 15

Ala Glu Gly Val Ser Glu Ser His Arg Ala Arg Ser Val Gly Ile Gly

20

25

30

Arg Ala His Ala Lys Ala Ile Leu Leu Gly Glu His Ala Val Val Tyr
 35 40 45

Gly Ala Pro Ala Leu Ala Leu Pro Ile Pro Gln Leu Thr Val Thr Ala
 50 55 60

Ser Val Gly Trp Ser Ser Glu Ala Ser Asp Ser Ala Gly Gly Leu Ser
 65 70 75 80

Tyr Thr Met Thr Gly Thr Pro Ser Arg Ala Leu Val Thr Gln Ala Ser
 85 90 95

Asp Gly Leu His Arg Leu Thr Ala Glu Phe Met Ala Arg Met Gly Val
 100 105 110

Thr Asn Ala Pro His Leu Asp Val Ile Leu Asp Gly Ala Ile Pro His
 115 120 125

Gly Arg Gly Leu Gly Ser Ser Ala Ala Gly Ser Arg Ala Ile Ala Leu
 130 135 140

Ala Leu Ala Asp Leu Phe Gly His Glu Leu Ala Glu His Thr Ala Tyr
 145 150 155 160

Glu Leu Val Gln Thr Ala Glu Asn Met Ala His Gly Arg Ala Ser Gly
 165 170 175

Val Asp Ala Met Thr Val Gly Ala Ser Arg Pro Leu Leu Phe Gln Gln
 180 185 190

Gly Arg Thr Glu Arg Leu Ala Ile Gly Cys Asp Ser Leu Phe Ile Val
 195 200 205

Ala Asp Ser Gly Val Pro Gly Ser Thr Lys Glu Ala Val Glu Met Leu
 210 215 220

Arg Glu Gly Phe Thr Arg Ser Ala Gly Thr Gln Glu Arg Phe Val Gly
 225 230 235 240

Arg Ala Thr Glu Leu Thr Glu Ala Ala Arg Gln Ala Leu Ala Asp Gly
 245 250 255

Arg Pro Glu Glu Leu Gly Ser Gln Leu Thr Tyr Tyr His Glu Leu Leu
 260 265 270

His Glu Ala Arg Leu Ser Thr Asp Gly Ile Asp Ala Leu Val Glu Ala
 275 280 285

Ala Leu Lys Ala Gly Ser Leu Gly Ala Lys Ile Thr Gly Gly Leu
 290 295 300

Gly Gly Cys Met Ile Ala Gln Ala Arg Pro Glu Gln Ala Arg Glu Val
 305 310 315 320

Thr Arg Gln Leu His Glu Ala Gly Ala Val Gln Thr Trp Val Val Pro

325

330

335

Leu Lys Gly Leu Asp Asn His Ala Gln
340 345

<210> 103

<211> 334

<212> PRT

<213> Streptomyces griseocephaloporeus

<400> 103

Met Thr Leu Pro Thr Ser Val Glu Glu Gly Ser Lys Ala His Arg Ala
1 5 10 15

Arg Ala Val Gly Thr Gly Arg Ala His Ala Lys Ala Ile Leu Leu Gly
20 25 30

Glu His Ala Val Val Tyr Gly Thr Pro Ala Leu Ala Met Pro Ile Pro
35 40 45

Gln Leu Ala Val Thr Ala Ser Ala Gly Trp Ser Gly Arg Ser Ala Glu
50 55 60

Ser Arg Gly Gly Pro Thr Phe Thr Met Thr Gly Ser Ala Ser Arg Ala
65 70 75 80

Val Thr Ala Gln Ala Leu Asp Gly Leu Arg Arg Leu Thr Ala Ser Val
85 90 95

Lys Ala His Thr Gly Val Thr Asp Gly Gln His Leu Asp Val Ser Leu
100 105 110

Asp Gly Ala Ile Pro Pro Gly Arg Gly Leu Gly Ser Ser Ala Ala Asn
115 120 125

Ala Arg Ala Ile Ile Leu Ala Leu Ala Asp Leu Phe Gly Arg Glu Leu
130 135 140

Thr Glu Gly Glu Val Phe Asp Leu Val Gln Glu Ala Glu Asn Leu Thr
145 150 155 160

His Gly Arg Ala Ser Gly Val Asp Ala Val Thr Val Gly Ala Thr Ala
165 170 175

Pro Leu Leu Phe Arg Ala Gly Thr Ala Gln Ala Leu Asp Ile Gly Cys
180 185 190

Asp Ala Leu Phe Val Val Ala Asp Ser Gly Thr Ala Gly Ser Thr Lys
195 200 205

Glu Ala Ile Glu Leu Leu Arg Ala Gly Phe Arg Ala Gly Ala Gly Lys

210	215	220
Glu Glu Arg Phe Met His Arg Ala Ala His Leu Val Asp Asp Ala Arg		
225	230	235
Ala Ser Leu Ala Glu Gly Glu Pro Glu Ala Phe Gly Ser Cys Leu Thr		
245	250	255
Glu Tyr His Gly Leu Leu Arg Gly Ala Gly Leu Ser Thr Asp Arg Ile		
260	265	270
Asp Ala Leu Val Asp Ala Ala Leu Gln Ala Asp Ser Leu Gly Ala Lys		
275	280	285
Ile Thr Gly Gly Leu Gly Gly Cys Val Leu Ala Met Ser Arg Pro		
290	295	300
Glu Arg Ala Glu Glu Val Ala Arg Gln Leu His Ala Ala Gly Ala Val		
305	310	315
320		
Arg Thr Trp Ala Val Gln Leu Arg Arg Ser Thr His Glu Arg		
325	330	
<210> 104		
<211> 296		
<212> PRT		
<213> <i>Borrelia burgdorferi</i>		
<400> 104		
Met Leu Arg Ile Arg Lys Pro Ala Lys Ile Leu Phe Leu Gly Glu His		
1	5	10
15		
Ser Ala Val Tyr Gly Phe Pro Val Ile Gly Ala Thr Val Pro Ile Tyr		
20	25	30
Met Asp Leu Ile Tyr Ser Val Ser Lys Asn Trp Lys Tyr Leu Gly Lys		
35	40	45
Pro Ser Thr Arg Leu Asn Ser Leu Ile Ser Phe Ile Val Ser Asn Tyr		
50	55	60
Ser Lys Val Asn Pro Ile Glu Phe Asp Ile Ile Ser Glu Ile Pro Ile		
65	70	75
80		
Gly Val Gly Leu Gly Ser Ser Ala Ser Leu Ser Leu Cys Phe Ala Glu		
85	90	95
Tyr Ile Thr Ser His Phe Glu Tyr Lys Asp Cys Asn Lys Ile Leu Leu		
100	105	110
Ala Asn Gln Ile Glu Asn Ile Phe His Gly Lys Ser Ser Gly Met Asp		

115	120	125
Ile Arg Leu Ile Asp Leu Asn Gly Thr Phe Tyr Leu Glu Lys Lys Glu		
130	135	140
Asn Val Leu His Ser Lys Lys Ile Lys Asp Ser Gly Phe Tyr Phe Leu		
145	150	155
Ile Gly Ala Ile Lys Arg Asp Leu Thr Thr Lys Glu Ile Val Val Asn		
165	170	175
Leu Lys Lys Asp Leu Leu Ser Asn Ala Tyr Leu Phe Val Phe Ile Glu		
180	185	190
Lys Leu Gly Leu Ala Val Ser Asn Ser Tyr Ala Ser Phe Gln Asn Lys		
195	200	205
Asp Val Tyr Ser Leu Ala Asn Glu Met Asn Ile Ala Gln Cys Cys Leu		
210	215	220
Lys Arg Leu Gly Leu Ser Asn Asp Thr Leu Asp Trp Leu Ile Ser Glu		
225	230	235
Gly Ile Lys Leu Gly Ala Leu Ser Gly Lys Leu Ser Gly Ala Gly Lys		
245	250	255
Gly Gly Ala Phe Ile Phe Leu Phe Glu Ser Leu Ile Lys Ala Asn Ile		
260	265	270
Val Gln Lys Glu Leu Asn Asn Met Leu Asp Ser Lys Ile Asp Leu Leu		
275	280	285
Leu Lys Leu Lys Val Ile Glu Thr		
290	295	
<210> 105		
<211> 336		
<212> PRT		
<213> Streptococcus pneumoniae		
<400> 105		
Met Ile Ala Val Lys Thr Cys Gly Lys Leu Tyr Trp Ala Gly Glu Tyr		
1	5	10
15		
Ala Ile Leu Glu Pro Gly Gln Leu Ala Leu Ile Lys Asp Ile Pro Ile		
20	25	30
Tyr Met Arg Ala Glu Ile Ala Phe Ser Asp Ser Tyr Arg Ile Tyr Ser		
35	40	45
Asp Met Phe Asp Phe Ala Val Asp Leu Arg Pro Asn Pro Asp Tyr Ser		

50	55	60
Leu Ile Gln Glu Thr Ile Ala Leu Met Gly Asp Phe Leu Ala Val Arg		
65	70	75
Gly Gln Asn Leu Arg Pro Phe Ser Leu Lys Ile Cys Gly Lys Met Glu		
85	90	95
Arg Glu Gly Lys Lys Phe Gly Leu Gly Ser Ser Gly Ser Val Val Val		
100	105	110
Leu Val Val Lys Ala Leu Leu Ala Leu Tyr Asn Leu Ser Val Asp Gln		
115	120	125
Asn Leu Leu Phe Lys Leu Thr Ser Ala Val Leu Leu Lys Arg Gly Asp		
130	135	140
Asn Gly Ser Met Gly Asp Leu Ala Cys Ile Val Ala Glu Asp Leu Val		
145	150	155
Leu Tyr Gln Ser Phe Asp Arg Gln Lys Ala Ala Ala Trp Leu Glu Glu		
165	170	175
Glu Asn Leu Ala Thr Val Leu Glu Arg Asp Trp Gly Phe Phe Ile Ser		
180	185	190
Gln Val Lys Pro Thr Leu Glu Cys Asp Phe Leu Val Gly Trp Thr Lys		
195	200	205
Glu Val Ala Val Ser Ser His Met Val Gln Gln Ile Lys Gln Asn Ile		
210	215	220
Asn Gln Asn Phe Leu Ser Ser Ser Lys Glu Thr Val Val Ser Leu Val		
225	230	235
Glu Ala Leu Glu Gln Gly Lys Ala Glu Lys Val Ile Glu Gln Val Glu		
245	250	255
Val Ala Ser Lys Leu Leu Glu Gly Leu Ser Thr Asp Ile Tyr Thr Pro		
260	265	270
Leu Leu Arg Gln Leu Lys Glu Ala Ser Gln Asp Leu Gln Ala Val Ala		
275	280	285
Lys Ser Ser Gly Ala Gly Gly Asp Cys Gly Ile Ala Leu Ser Phe		
290	295	300
Asp Ala Gln Ser Ser Arg Asn Thr Leu Lys Asn Arg Trp Ala Asp Leu		
305	310	315
Gly Ile Glu Leu Leu Tyr Gln Glu Arg Ile Gly His Asp Asp Lys Ser		
325	330	335
<210> 106		
<211> 335		

<212> PRT

<213> Streptococcus pyogenes

<400> 106

Met Ser Asn Tyr Cys Val Gln Thr Gly Gly Lys Leu Tyr Leu Thr Gly
1 5 10 15

Glu Tyr Ala Ile Leu Ile Pro Gly Gln Lys Ala Leu Ile His Phe Ile
20 25 30

Pro Leu Met Met Thr Ala Glu Ile Ser Pro Ala Ala His Ile Gln Leu
35 40 45

Ala Ser Asp Met Phe Ser His Lys Ala Gly Met Thr Pro Asp Ala Ser
50 55 60

Tyr Ala Leu Ile Gln Ala Thr Val Lys Thr Phe Ala Asp Tyr Leu Gly
65 70 75 80

Gln Ser Ile Asp Gln Leu Glu Pro Phe Ser Leu Ile Ile Thr Gly Lys
85 90 95

Met Glu Arg Asp Gly Lys Lys Phe Gly Ile Gly Ser Ser Gly Ser Val
100 105 110

Thr Leu Leu Thr Leu Lys Ala Leu Ser Ala Tyr Tyr Gln Ile Thr Leu
115 120 125

Thr Pro Glu Leu Leu Phe Lys Leu Ala Ala Tyr Thr Leu Leu Lys Gln
130 135 140

Gly Asp Asn Gly Ser Met Gly Asp Ile Ala Cys Ile Ala Tyr Gln Thr
145 150 155 160

Leu Val Ala Tyr Thr Ser Phe Asp Arg Glu Gln Val Ser Asn Trp Leu
165 170 175

Gln Thr Met Pro Leu Lys Lys Leu Val Lys Asp Trp Gly Tyr His
180 185 190

Ile Gln Val Ile Gln Pro Ala Leu Pro Cys Asp Phe Leu Val Gly Trp
195 200 205

Thr Lys Ile Pro Ala Ile Ser Arg Gln Met Ile Gln Gln Val Thr Ala
210 215 220

Ser Ile Thr Pro Ala Phe Leu Arg Thr Ser Tyr Gln Leu Thr Gln Ser
225 230 235 240

Ala Met Val Ala Leu Gln Glu Gly His Lys Glu Glu Leu Lys Lys Ser
245 250 255

Leu Ala Gly Ala Ser His Leu Leu Lys Glu Leu His Pro Ala Ile Tyr

260

265

270

His Pro Lys Leu Val Thr Leu Val Ala Ala Cys Gln Lys Gln Asp Ala
 275 280 285

Val Ala Lys Ser Ser Gly Ala Gly Gly Asp Cys Gly Ile Ala Leu
 290 295 300

Ala Phe Asn Gln Asp Ala Arg Asp Thr Leu Ile Ser Lys Trp Gln Glu
 305 310 315 320

Ala Asp Ile Ala Leu Leu Tyr Gln Glu Arg Trp Gly Glu Asn Asp
 325 330 335

<210> 107

<211> 368

<212> PRT

<213> Enterococcus faecalis

<400> 107

Met Ile Glu Val Thr Thr Pro Gly Lys Leu Phe Ile Ala Gly Glu Tyr
 1 5 10 15

Ala Val Val Glu Pro Gly His Pro Ala Ile Ile Val Ala Val Asp Gln
 20 25 30

Phe Val Thr Val Thr Val Glu Glu Thr Thr Asp Glu Gly Ser Ile Gln
 35 40 45

Ser Ala Gln Tyr Ser Ser Leu Pro Ile Arg Trp Thr Arg Arg Asn Gly
 50 55 60

Glu Leu Val Leu Asp Ile Arg Glu Asn Pro Phe His Tyr Val Leu Ala
 65 70 75 80

Ala Ile His Leu Thr Glu Lys Tyr Ala Gln Glu Gln Asn Lys Glu Leu
 85 90 95

Ser Phe Tyr His Leu Lys Val Thr Ser Glu Leu Asp Ser Ser Asn Gly
 100 105 110

Arg Lys Tyr Gly Leu Gly Ser Ser Gly Ala Val Thr Val Gly Thr Val
 115 120 125

Lys Ala Leu Asn Ile Phe Tyr Asp Leu Gly Leu Glu Asn Glu Glu Ile
 130 135 140

Phe Lys Leu Ser Ala Leu Ala His Leu Ala Val Gln Gly Asn Gly Ser
 145 150 155 160

Cys Gly Asp Ile Ala Ala Ser Cys Tyr Gly Trp Ile Ala Phe Ser

	165	170	175
Thr Phe Asp His Asp Trp Val Asn Gln Lys Val Thr Thr Glu Thr Leu			
180	185		190
Thr Asp Leu Leu Ala Met Asp Trp Pro Glu Leu Met Ile Phe Pro Leu			
195	200		205
Lys Val Pro Lys Gln Leu Arg Leu Leu Ile Gly Trp Thr Gly Ser Pro			
210	215		220
Ala Ser Thr Ser Asp Leu Val Asp Arg Val His Gln Ser Lys Glu Glu			
225	230	235	240
Lys Gln Ala Ala Tyr Glu Gln Phe Leu Met Lys Ser Arg Leu Cys Val			
245	250		255
Glu Thr Met Ile Asn Gly Phe Asn Thr Gly Lys Ile Ser Val Ile Gln			
260	265		270
Lys Gln Ile Thr Lys Asn Arg Gln Leu Leu Ala Glu Leu Ser Ser Leu			
275	280		285
Thr Gly Val Val Ile Glu Thr Glu Ala Leu Lys Asn Leu Cys Asp Leu			
290	295		300
Ala Glu Ser Tyr Thr Gly Ala Ala Lys Ser Ser Gly Ala Gly Gly			
305	310	315	320
Asp Cys Gly Ile Val Ile Phe Arg Gln Lys Ser Gly Ile Leu Pro Leu			
325	330		335
Met Thr Ala Trp Glu Lys Asp Gly Ile Thr Pro Leu Pro Leu His Val			
340	345		350
Tyr Thr Tyr Gly Gln Lys Glu Cys Lys Glu Lys His Glu Ser Lys Arg			
355	360		365
<210> 108			
<211> 361			
<212> PRT			
<213> Enterococcus faecium			
<400> 108			
Met Ile Glu Val Ser Ala Pro Gly Lys Leu Tyr Ile Ala Gly Glu Tyr			
1	5	10	15
Ala Val Val Glu Thr Gly His Pro Ala Val Ile Ala Ala Val Asp Gln			
20	25		30
Phe Val Thr Val Thr Val Glu Ser Ala Arg Lys Val Gly Ser Ile Gln			

35

40

45

Ser Ala Gln Tyr Ser Gly Met Pro Val Arg Trp Thr Arg Arg Asn Gly
 50 55 60

Glu Leu Val Leu Asp Ile Arg Glu Asn Pro Phe His Tyr Ile Leu Ala
 65 70 75 80

Ala Ile Arg Leu Thr Glu Lys Tyr Ala Gln Glu Lys Asn Ile Leu Leu
 85 90 95

Ser Phe Tyr Asp Leu Lys Val Thr Ser Glu Leu Asp Ser Ser Asn Gly
 100 105 110

Arg Lys Tyr Gly Leu Gly Ser Ser Gly Ala Val Thr Val Ala Thr Val
 115 120 125

Lys Ala Leu Asn Val Phe Tyr Ala Leu Asn Leu Ser Gln Leu Glu Ile
 130 135 140

Phe Lys Ile Ala Ala Leu Ala Asn Leu Ala Val Gln Asp Asn Gly Ser
 145 150 155 160

Cys Gly Asp Ile Ala Ala Ser Cys Tyr Gly Gly Trp Ile Ala Phe Ser
 165 170 175

Thr Phe Asp His Pro Trp Leu Gln Glu Gln Glu Thr Gln His Ser Ile
 180 185 190

Ser Glu Leu Leu Ala Leu Asp Trp Pro Gly Leu Ser Ile Glu Pro Leu
 195 200 205

Ile Ala Pro Glu Asp Leu Arg Leu Leu Ile Gly Trp Thr Gly Ser Pro
 210 215 220

Ala Ser Thr Ser Asp Leu Val Asp Gln Val His Arg Ser Arg Glu Asp
 225 230 235 240

Lys Met Val Ala Tyr Gln Leu Phe Leu Lys Asn Ser Thr Glu Cys Val
 245 250 255

Asn Glu Met Ile Lys Gly Phe Lys Glu Asn Asn Val Thr Leu Ile Gln
 260 265 270

Gln Met Ile Arg Lys Asn Arg Gln Leu Leu His Asp Leu Ser Ala Ile
 275 280 285

Thr Gly Val Val Ile Glu Thr Pro Ala Leu Asn Lys Leu Cys Asn Leu
 290 295 300

Ala Glu Gln Tyr Glu Gly Ala Ala Lys Ser Ser Gly Ala Gly Gly Gly
 305 310 315 320

Asp Cys Gly Ile Val Ile Val Asp Gln Lys Ser Gly Ile Leu Pro Leu
 325 330 335

Met Ser Ala Trp Glu Lys Ala Glu Ile Thr Pro Leu Pro Leu His Val

340

345

350

Tyr Ser Asp Gln Arg Lys Glu Asn Arg
 355 360

<210> 109

<211> 358

<212> PRT

<213> Staphylococcus haemolyticus

<400> 109

Met Ile Gln Val Lys Ala Pro Gly Lys Leu Tyr Val Ala Gly Glu Tyr
 1 5 10 15

Ala Val Thr Glu Pro Gly Tyr Lys Ser Val Leu Ile Ala Val Asp Arg
 20 25 30

Phe Val Thr Ala Ser Ile Glu Ala Ser Asn Ala Val Thr Ser Thr Ile
 35 40 45

His Ser Lys Thr Leu His Tyr Glu Pro Val Thr Phe Asn Arg Asn Glu
 50 55 60

Asp Lys Ile Asp Ile Ser Asp Ala Asn Ala Ala Ser Gln Leu Lys Tyr
 65 70 75 80

Val Val Thr Ala Ile Glu Val Phe Glu Gln Tyr Ala Arg Ser Cys Asn
 85 90 95

Val Lys Leu Lys His Phe His Leu Glu Ile Asp Ser Asn Leu Asp Asp
 100 105 110

Ala Ser Gly Asn Lys Tyr Gly Leu Gly Ser Ser Ala Ala Val Leu Val
 115 120 125

Ser Val Val Lys Ala Leu Asn Glu Phe Tyr Asp Met Gln Leu Ser Asn
 130 135 140

Leu Tyr Ile Tyr Lys Leu Ala Val Ile Ser Asn Met Arg Leu Gln Ser
 145 150 155 160

Leu Ser Ser Cys Gly Asp Ile Ala Val Ser Val Tyr Ser Gly Trp Leu
 165 170 175

Ala Tyr Ser Thr Phe Asp His Asp Trp Val Lys Gln Gln Met Glu Glu
 180 185 190

Thr Ser Val Asn Glu Val Leu Glu Lys Asn Trp Pro Gly Leu His Ile
 195 200 205

Glu Pro Leu Gln Ala Pro Glu Asn Met Glu Val Leu Ile Gly Trp Thr

210	215	220
Gly Ser Pro Ala Ser Ser Pro His Leu Val Ser Glu Val Lys Arg Leu		
225	230	235
Lys Ser Asp Pro Ser Phe Tyr Gly Arg Phe Leu Asp Gln Ser His Thr		
245	250	255
Cys Val Glu Asn Leu Ile Tyr Ala Phe Lys Thr Asp Asn Ile Lys Gly		
260	265	270
Val Gln Lys Met Ile Arg Gln Asn Arg Met Ile Ile Gln Gln Met Asp		
275	280	285
Asn Glu Ala Thr Val Asp Ile Glu Thr Glu Asn Leu Lys Met Leu Cys		
290	295	300
Asp Ile Gly Glu Arg Tyr Gly Ala Ala Ala Lys Thr Ser Gly Ala Gly		
305	310	315
Gly Gly Asp Cys Gly Ile Ala Ile Ile Asp Asn Arg Ile Asp Lys Asn		
325	330	335
Arg Ile Tyr Asn Glu Trp Ala Ser His Gly Ile Lys Pro Leu Lys Phe		
340	345	350
Lys Ile Tyr His Gly Gln		
355		
<210> 110		
<211> 358		
<212> PRT		
<213> Staphylococcus epidermidis		
<400> 110		
Met Ile Gln Val Lys Ala Pro Gly Lys Leu Tyr Ile Ala Gly Glu Tyr		
1	5	10
		15
Ala Val Thr Glu Pro Gly Tyr Lys Ser Ile Leu Ile Ala Val Asn Arg		
20	25	30
Phe Val Thr Ala Thr Ile Glu Ala Ser Asn Lys Val Glu Gly Ser Ile		
35	40	45
His Ser Lys Thr Leu His Tyr Glu Pro Val Lys Phe Asp Arg Asn Glu		
50	55	60
Asp Arg Ile Glu Ile Ser Asp Val Gln Ala Ala Lys Gln Leu Lys Tyr		
65	70	75
		80
Val Val Thr Ala Ile Glu Val Phe Glu Gln Tyr Val Arg Ser Cys Asn		

85	90	95
Met Asn Leu Lys His Phe His	Leu Thr Ile Asp Ser Asn	Leu Ala Asp
100	105	110
Asn Ser Gly Gln Lys Tyr Gly	Leu Gly Ser Ser	Ala Ala Val Leu Val
115	120	125
Ser Val Val Lys Ala Leu Asn Glu Phe Tyr Gly	Leu Glu Leu Ser Asn	
130	135	140
Leu Tyr Ile Tyr Lys Leu Ala Val Ile Ala Asn Met Lys Leu Gln Ser		
145	150	155
Leu Ser Ser Cys Gly Asp Ile Ala Val Ser Val Tyr Ser Gly Trp Leu		
165	170	175
Ala Tyr Ser Thr Phe Asp His Asp Trp Val Lys Gln Gln Met Glu Glu		
180	185	190
Thr Ser Val Asn Asp Val Leu Glu Lys Asn Trp Pro Gly Leu His Ile		
195	200	205
Glu Pro Leu Gln Ala Pro Glu Asn Met Glu Val Leu Ile Gly Trp Thr		
210	215	220
Gly Ser Pro Ala Ser Ser Pro His Leu Val Ser Glu Val Lys Arg Leu		
225	230	235
240		
Lys Ser Asp Pro Ser Phe Tyr Gly Asp Phe Leu Asp Gln Ser His Ala		
245	250	255
Cys Val Glu Ser Leu Ile Gln Ala Phe Lys Thr Asn Asn Ile Lys Gly		
260	265	270
Val Gln Lys Met Ile Arg Ile Asn Arg Arg Ile Ile Gln Ser Met Asp		
275	280	285
Asn Glu Ala Ser Val Glu Ile Glu Thr Asp Lys Leu Lys Lys Leu Cys		
290	295	300
Asp Val Gly Glu Lys His Gly Gly Ala Ser Lys Thr Ser Gly Ala Gly		
305	310	315
320		
Gly Gly Asp Cys Gly Ile Thr Ile Asn Lys Val Ile Asp Lys Asn		
325	330	335
Ile Ile Tyr Asn Glu Trp Gln Met Asn Asp Ile Lys Pro Leu Lys Phe		
340	345	350
Lys Ile Tyr His Gly Gln		
355		
<210> 111		
<211> 358		

<212> PRT

<213> Staphylococcus aureus

<400> 111

Met Ile Gln Val Lys Ala Pro Gly Lys Leu Tyr Ile Ala Gly Glu Tyr
1 5 10 15

Ala Val Thr Glu Pro Gly Tyr Lys Ser Val Leu Ile Ala Leu Asp Arg
20 25 30

Phe Val Thr Ala Thr Ile Glu Glu Ala Thr Gln Tyr Lys Gly Thr Ile
35 40 45

His Ser Lys Ala Leu His His Asn Pro Val Thr Phe Ser Arg Asp Glu
50 55 60

Asp Ser Ile Val Ile Ser Asp Pro His Ala Ala Lys Gln Leu Asn Tyr
65 70 75 80

Val Val Thr Ala Ile Glu Ile Phe Glu Gln Tyr Ala Lys Ser Cys Asp
85 90 95

Ile Ala Met Lys His Phe His Leu Thr Ile Asp Ser Asn Leu Asp Asp
100 105 110

Ser Asn Gly His Lys Tyr Gly Leu Gly Ser Ser Ala Ala Val Leu Val
115 120 125

Ser Val Ile Lys Val Leu Asn Glu Phe Tyr Asp Met Lys Leu Ser Asn
130 135 140

Leu Tyr Ile Tyr Lys Leu Ala Val Ile Ala Asn Met Lys Leu Gln Ser
145 150 155 160

Leu Ser Ser Cys Gly Asp Ile Ala Val Ser Val Tyr Ser Gly Trp Leu
165 170 175

Ala Tyr Ser Thr Phe Asp His Glu Trp Val Lys His Gln Ile Glu Asp
180 185 190

Thr Thr Val Glu Glu Val Leu Ile Lys Asn Trp Pro Gly Leu His Ile
195 200 205

Glu Pro Leu Gln Ala Pro Glu Asn Met Glu Val Leu Ile Gly Trp Thr
210 215 220

Gly Ser Pro Ala Ser Ser Pro His Phe Val Ser Glu Val Lys Arg Leu
225 230 235 240

Lys Ser Asp Pro Ser Phe Tyr Gly Asp Phe Leu Glu Asp Ser His Arg
245 250 255

Cys Val Glu Lys Leu Ile His Ala Phe Lys Thr Asn Asn Ile Lys Gly

260

265

270

Val Gln Lys Met Val Arg Gln Asn Arg Thr Ile Ile Gln Arg Met Asp
 275 280 285

Lys Glu Ala Thr Val Asp Ile Glu Thr Glu Lys Leu Lys Tyr Leu Cys
 290 295 300

Asp Ile Ala Glu Lys Tyr His Gly Ala Ser Lys Thr Ser Gly Ala Gly
 305 310 315 320

Gly Gly Asp Cys Gly Ile Thr Ile Ile Asn Lys Asp Val Asp Lys Glu
 325 330 335

Lys Ile Tyr Asp Glu Trp Thr Lys His Gly Ile Lys Pro Leu Lys Phe
 340 345 350

Asn Ile Tyr His Gly Gln
 355

<210> 112

<211> 374

<212> PRT

<213> Streptomyces sp. CL190

<400> 112

Met Thr Thr Gly Gln Arg Thr Ile Val Arg His Ala Pro Gly Lys Leu
 1 5 10 15

Phe Val Ala Gly Glu Tyr Ala Val Val Asp Pro Gly Asn Pro Ala Ile
 20 25 30

Leu Val Ala Val Asp Arg His Ile Ser Val Thr Val Ser Asp Ala Asp
 35 40 45

Ala Asp Thr Gly Ala Ala Asp Val Val Ile Ser Ser Asp Leu Gly Pro
 50 55 60

Gln Ala Val Gly Trp Arg Trp His Asp Gly Arg Leu Val Val Arg Asp
 65 70 75 80

Pro Asp Asp Gly Gln Gln Ala Arg Ser Ala Leu Ala His Val Val Ser
 85 90 95

Ala Ile Glu Thr Val Gly Arg Leu Leu Gly Glu Arg Gly Gln Lys Val
 100 105 110

Pro Ala Leu Thr Leu Ser Val Ser Ser Arg Leu His Glu Asp Gly Arg
 115 120 125

Lys Phe Gly Leu Gly Ser Ser Gly Ala Val Thr Val Ala Thr Val Ala

130	135	140
Ala Val Ala Ala Phe Cys Gly Leu Glu Leu Ser Thr Asp Glu Arg Phe		
145	150	155
Arg Leu Ala Met Leu Ala Thr Ala Glu Leu Asp Pro Lys Gly Ser Gly		
165	170	175
Gly Asp Leu Ala Ala Ser Thr Trp Gly Gly Trp Ile Ala Tyr Gln Ala		
180	185	190
Pro Asp Arg Ala Phe Val Leu Asp Leu Ala Arg Arg Val Gly Val Asp		
195	200	205
Arg Thr Leu Lys Ala Pro Trp Pro Gly His Ser Val Arg Arg Leu Pro		
210	215	220
Ala Pro Lys Gly Leu Thr Leu Glu Val Gly Trp Thr Gly Glu Pro Ala		
225	230	235
Ser Thr Ala Ser Leu Val Ser Asp Leu His Arg Arg Thr Trp Arg Gly		
245	250	255
Ser Ala Ser His Gln Arg Phe Val Glu Thr Thr Asp Cys Val Arg		
260	265	270
Ser Ala Val Thr Ala Leu Glu Ser Gly Asp Asp Thr Ser Leu Leu His		
275	280	285
Glu Ile Arg Arg Ala Arg Gln Glu Leu Ala Arg Leu Asp Asp Glu Val		
290	295	300
Gly Leu Gly Ile Phe Thr Pro Lys Leu Thr Ala Leu Cys Asp Ala Ala		
305	310	315
Glu Ala Val Gly Gly Ala Ala Lys Pro Ser Gly Ala Gly Gly Asp		
325	330	335
Cys Gly Ile Ala Leu Leu Asp Ala Glu Ala Ser Arg Asp Ile Thr His		
340	345	350
Val Arg Gln Arg Trp Glu Thr Ala Gly Val Leu Pro Leu Pro Leu Thr		
355	360	
Pro Ala Leu Glu Gly Ile		
370		
<210> 113		
<211> 360		
<212> PRT		
<213> Streptomyces griseolosporeus		

<400> 113

Met Thr Gly Pro Arg Ala Val Thr Arg Arg Ala Pro Gly Lys Leu Phe
1 5 10 15

Val Ala Gly Glu Tyr Ala Val Val Glu Pro Gly Asn Arg Ala Ile Leu
20 25 30

Val Ala Val Asp Arg Tyr Val Thr Val Thr Val Ser Asp Gly Ala Ala
35 40 45

Pro Gly Val Val Val Ser Ser Asp Ile Gly Ala Gly Pro Val His His
50 55 60

Pro Trp Gln Asp Gly Arg Leu Thr Gly Gly Thr Gly Thr Pro His Val
65 70 75 80

Val Ala Ala Val Glu Thr Val Ala Arg Leu Leu Ala Glu Arg Gly Arg
85 90 95

Ser Val Pro Pro Leu Gly Trp Ser Ile Ser Ser Thr Leu His Glu Asp
100 105 110

Gly Arg Lys Phe Gly Leu Gly Ser Ser Gly Ala Val Thr Val Ala Thr
115 120 125

Val Ser Ala Val Ala Ala His Cys Gly Leu Glu Leu Thr Ala Glu Glu
130 135 140

Arg Phe Arg Thr Ala Leu Ile Ala Ser Ala Arg Ile Asp Pro Arg Gly
145 150 155 160

Ser Gly Gly Asp Ile Ala Thr Ser Thr Trp Gly Gly Trp Ile Ala Tyr
165 170 175

Arg Ala Pro Asp Arg Asp Ala Val Leu Asp Leu Thr Arg Arg Gln Gly
180 185 190

Val Asp Glu Ala Leu Arg Ala Pro Trp Pro Gly Phe Ser Val Arg Leu
195 200 205

Ser Pro Pro Arg Asn Leu Cys Leu Glu Val Gly Trp Thr Gly Asn Pro
210 215 220

Val Ser Thr Thr Ser Leu Leu Thr Asp Leu His Arg Arg Thr Trp Arg
225 230 235 240

Gly Ser Pro Ala Tyr Arg Arg Tyr Val Gly Ala Thr Gly Glu Leu Val
245 250 255

Asp Ala Ala Val Ile Ala Leu Glu Asp Gly Asp Thr Glu Gly Leu Leu
260 265 270

Arg Gln Val Arg Arg Ala Arg His Glu Met Val Arg Leu Asp Asp Glu
275 280 285

Val Gly Leu Gly Ile Phe Thr Pro Glu Leu Thr Ala Leu Cys Ala Ile

290

295

300

Ala Glu Arg Ala Gly Ala Ala Lys Pro Ser Gly Ala Gly Gly Gly Asp
 305 310 315 320

Cys Gly Ile Ala Leu Leu Asp Ala Glu Ala Arg Tyr Asp Arg Ser Pro
 325 330 335

Leu His Arg Gln Trp Ala Ala Ala Gly Val Leu Pro Leu Leu Val Ser
 340 345 350

Pro Ala Thr Glu Gly Val Glu Glu
 355 360

<210> 114

<211> 317

<212> PRT

<213> *Borrelia burgdorferi*

<400> 114

Met Asp Leu Ile Ser Phe Ser Val Pro Gly Asn Leu Leu Leu Met Gly
 1 5 10 15

Glu Tyr Thr Ile Leu Glu Glu Lys Gly Leu Gly Leu Ala Ile Ala Ile
 20 25 30

Asn Lys Arg Ala Phe Phe Ser Phe Lys Lys Ser Asp Ser Trp Arg Phe
 35 40 45

Phe Ser Lys Lys Lys Ile Asp Asp Phe Ser Leu Ile Glu Asn Arg
 50 55 60

Ser Asp Phe Val Phe Lys Met Phe Ala Tyr Leu Ser Gln Asn Cys Phe
 65 70 75 80

Phe Asn Leu Glu Asn Phe Ala Tyr Asp Val Tyr Ile Asp Thr Ser Asn
 85 90 95

Phe Phe Phe Asn Asp Gly Thr Lys Lys Gly Phe Gly Ser Ser Ala Val
 100 105 110

Val Ala Ile Gly Ile Val Cys Gly Leu Phe Leu Ile His Asn Ala Thr
 115 120 125

Asn Val Val Glu Lys Gly Glu Ile Phe Lys Tyr Cys Leu Glu Ala Tyr
 130 135 140

Arg Tyr Ser Gln Gly Gly Ile Gly Ser Gly Tyr Asp Ile Ala Thr Ser
 145 150 155 160

Ile Phe Gly Gly Val Ile Glu Phe Glu Gly Phe Asn Pro Lys Cys

165 170 175

Arg Gln Leu Gly Ala Val Glu Phe Asn Asp Phe Tyr Leu Met Gln Gly
180 185 190

Leu Gln Ala Ile Lys Thr Thr Ser Ile Cys Glu Tyr Asn Lys His
195 200 205

Arg Asn Ser Ile Leu Asp Phe Ile Leu Lys Cys Asn Leu Glu Met Lys
210 215 220

Lys Leu Val Leu Asn Ala Ser Asn Ser Lys Ser Ala Leu Ile Ser Ser
225 230 235 240

Leu Arg Arg Ala Lys Glu Leu Gly Leu Ala Ile Gly Glu Ala Ile Gly
245 250 255

Val Ser Ala Ala Leu Pro Ser Ser Phe Asp His Leu Leu Gly Gln Cys
260 265 270

Asp Leu Ile Lys Ala Leu Gly Ala Gly Asn Glu Thr Phe Leu Val Tyr
275 280 285

Arg Pro Asn Ile Glu Ala Phe Asn Leu Ser Lys Ile Ile Ser Ile Val
290 295 300

Leu Glu Asn Glu Gly Ile Lys Phe Glu Ser Asp Lys Cys
305 310 315

<210> 115

<211> 30

<212> DNA

<213> synthetic construct

<400> 115
ggcaagctt gtccacggca cgaccaagca

30

<210> 116

<211> 30

<212> DNA

<213> synthetic construct

<400> 116
cgtaatccgc ggccgcgttt ccagcgcgtc

30

<210> 117
<211> 28
<212> DNA
<213> synthetic construct

<400> 117
aattaaagga gggtttcata tgaattcg

28

<210> 118
<211> 28
<212> DNA
<213> synthetic construct

<400> 118
gatccgaatt catatgaaac cctccttt

28

<210> 119
<211> 30
<212> DNA
<213> synthetic construct

<400> 119
aaggcctcat atgattccc ataccccggt

30

<210> 120
<211> 28
<212> DNA
<213> synthetic construct

<400> 120
cgggatcctc atcgctccat ctccatgt

28

<210> 121

<211> 30

<212> DNA

<213> synthetic construct

<400> 121
aaggcctcat atgaccgaca gcaaggatca

30

<210> 122

<211> 28

<212> DNA

<213> synthetic construct

<400> 122
cgggatcctc attgacggat aagcgagg

28

<210> 123

<211> 29

<212> DNA

<213> synthetic construct

<400> 123
aaggcctcat atgaaagtgc ctaagatga

29

<210> 124

<211> 28

<212> DNA

<213> synthetic construct

<400> 124
cgggatcctc aggctgccg gtcgacat

28

<210> 125

<211> 34

<212> DNA

<213> synthetic construct

<400> 125

aaggcctcat atgagcacccg gcaggcctga agca

34

<210> 126

<211> 31

<212> DNA

<213> synthetic construct

<400> 126

cgggatcctc atccctgccc cggcagcggt t

31

<210> 127

<211> 30

<212> DNA

<213> synthetic construct

<400> 127

aaggcctcat atggatcagg tcatccgcgc

30

<210> 128

<211> 28

<212> DNA

<213> synthetic construct

<400> 128

cgggatcctc agtcatcgaa aacaagtc

28

<210> 129

<211> 30

<212> DNA

<213> synthetic construct

<400> 129

aaggcctcat atgactgatg ccgtccgcga

30

<210> 130

<211> 28

<212> DNA

<213> synthetic construct

<400> 130

cgggatcctc aacgcccctc gaacggcg

28

<210> 131

<211> 31

<212> DNA

<213> synthetic construct

<400> 131

ccggcattcg ggcggcatcc aggttcgcgt g

31

<210> 132

<211> 31

<212> DNA

<213> synthetic construct

<400> 132

cagcgagacc tggatgccgc ccgaatgccg g

31

<210> 133

<211> 31

<212> DNA

<213> synthetic construct

<400> 133

cgtgcaggc tggattctgt cggaataccc g

31

<210> 134

<211> 31

<212> DNA

<213> synthetic construct

<400> 134

cgggtattcc gacagaatcc agccctgcac g

31

<210> 135

<211> 31

<212> DNA

<213> synthetic construct

<400> 135

gggctgcgcg ccggcatccg gcatttcgac g

31

<210> 136

<211> 31

<212> DNA

<213> synthetic construct

<400> 136

cgtcgaaatg ccggatgccg gcgcgcagcc c

31

<210> 137

<211> 31

<212> DNA

<213> synthetic construct

<400> 137
gggtgcgacg ggcgagttct tcgatgcgca g

31

<210> 138

<211> 31

<212> DNA

<213> synthetic construct

<400> 138
ccgcgcacatcg aagaactcgc ccgtcgacc c

31

<210> 139

<211> 31

<212> DNA

<213> synthetic construct

<400> 139
cacgccccgtc acatacgacg aatacgttgc c

31

<210> 140

<211> 31

<212> DNA

<213> synthetic construct

<400> 140
ggcaacgtat tcgtcgatgc tgacgggcgt g

31

<210> 141

<211> 31

<212> DNA

<213> synthetic construct

<400> 141

gaggctcggg cttggctcct cggcggcggt g

31

<210> 142

<211> 31

<212> DNA

<213> synthetic construct

(C) <400> 142

caccgcggcc gaggagccaa gcccgagcct c

31

<210> 143

<211> 31

<212> DNA

<213> synthetic construct

<400> 143

cggcacgctg ctggacccgg gcgacgcctt c

31

<210> 144

<211> 31

<212> DNA

<213> synthetic construct

<400> 144

gaaggcgtcg cccgggtcca gcagcgtgcc g

31

<210> 145

<211> 36

<212> DNA

<213> synthetic construct

<400> 145
tcagaattcg gtaccatatg aagcttggat ccgggg

36

<210> 146

<211> 29

<212> DNA

<213> synthetic construct

<400> 146
ggatccaaaggc ttcatatggt accgaattc

29

<210> 147

<211> 26

<212> DNA

<213> synthetic construct

<400> 147
ggaattcgct gctgaacgcg atggcg

26

<210> 148

<211> 32

<212> DNA

<213> synthetic construct

<400> 148
ggggtaaccat atgtgccttc gttgcgtcag tc

32

<210> 149

<211> 50

<212> DNA

<213> synthetic construct

10/560726

WPI8 Rec'd PCT/PTO 08 DEC 2005

<400> 149
gatccggcgt gtgcgcaatt taattgcgca cacgccccct gcgtttaaac 50

<210> 150

<211> 50

<212> DNA

<213> synthetic construct

<400> 150
gatcgtttaa acgcaggggg cgtgtgcgca attaaattgc gcacacgccc 50

<210> 151

<211> 30

<212> DNA

<213> synthetic construct

<400> 151
aaggcctcat atgacgccc a agcagcaatt 30

<210> 152

<211> 26

<212> DNA

<213> synthetic construct

<400> 152
cgggatccta ggcgctgcgg cggatg 26

<210> 153

<211> 30

<212> DNA

<213> synthetic construct

<400> 153
ccggatcctc atgcctgccc gtcgacata 30
<210> 154
<211> 30
<212> DNA
<213> synthetic construct

<400> 154
gaaggcacat atgaatcagg tcatccgcgc 30
<210> 155
<211> 30
<212> DNA
<213> synthetic construct

<400> 155
gccggatcct cattcatcga aaacaagtcc 30
<210> 156
<211> 30
<212> DNA
<213> synthetic construct

<400> 156
acgcccggatc ctcatcgccc ctcgaacggc 30
<210> 157
<211> 1612
<212> DNA
<213> Paracoccus sp. R114

<220>

<221> CDS

<222> (59)..(292)

<223> XseB

<220>

<221> CDS

<222> (1185)..(1610)

<223> IspA

<220>

<221> CDS

<222> (295)..(1158)

<223> Dxs

<400> 157

ccatggcatt cgggtcgat ggcgtctatg ttggcccgaa caggcagcag gaggcccc 58

atg	agc	gat	atc	cag	acc	ctc	tcg	tcc	gag	gaa	gcc	atg	cgc	gag	ctg
Met	Ser	Asp	Ile	Gln	Thr	Leu	Ser	Phe	Glu	Glu	Ala	Met	Arg	Glu	Leu
1			5									10			15

gag	gca	acc	gtc	ggc	aag	ctg	gaa	acc	ggc	gag	gca	acg	ctc	gag	gac
Glu	Ala	Thr	Val	Gly	Lys	Leu	Glu	Thr	Gly	Glu	Ala	Thr	Leu	Glu	Asp
20			25									30			

tcc	atc	gca	ctc	tat	gaa	cgc	ggg	gca	gca	ctg	cgc	gcc	cat	tgc	gaa
Ser	Ile	Ala	Leu	Tyr	Glu	Arg	Gly	Ala	Ala	Leu	Arg	Ala	His	Cys	Glu
35			40									45			

acc	cgc	ctg	cgc	gag	gcc	gag	cgg	gtc	gag	aag	atc	acc	ctg	gcc	
Thr	Arg	Leu	Arg	Glu	Ala	Glu	Glu	Arg	Val	Glu	Lys	Ile	Thr	Leu	Ala
50			55								60				

gca	aac	ggg	cag	ccg	tcc	gga	acc	gag	ccc	gcc	gag	ggc	ctg	tg	atg
Ala	Asn	Gly	Gln	Pro	Ser	Gly	Thr	Glu	Pro	Ala	Glu	Gly	Leu		Met
65	*		70										*		

cag	gcc	cgc	ctg	gcc	gag	atc	cgg	ccc	ctg	gtc	gag	gcc	gag	ctg	aac
Gln	Ala	Arg	Leu	Ala	Glu	Ile	Arg	Pro	Leu	Val	Glu	Ala	Glu	Leu	Asn
80			85									90		95	

gcc gcc atc gac gcg ctg ccc ggc gat ctg tcg gat gcg atg cgc Ala Ala Ile Asp Ala Leu Pro Ala Gly Asp Leu Ser Asp Ala Met Arg 100 105 110	393
tat gcc gtg cag ggc ggc aag cgg ctg cgc gcg ttc ctg gtg atg gag Tyr Ala Val Gln Gly Gly Lys Arg Leu Arg Ala Phe Leu Val Met Glu 115 120 125	441
tcg gcg cgc ctg cac ggg ctg gac gac gca tcg ctg ccc gtc gcc Ser Ala Arg Leu His Gly Leu Asp Asp Ala Ser Leu Pro Val Ala 130 135 140	489
gcc gcg gtc gag gcg ctg cac gac tac agc ttg gtc cat gac gac ctg Ala Ala Val Glu Ala Leu His Ala Tyr Ser Leu Val His Asp Asp Leu 145 150 155	537
ccc gcg atg gat gac gac ctg cgg cgc ggt cag ccc acc gtc cac Pro Ala Met Asp Asp Asp Leu Arg Arg Gly Gln Pro Thr Val His 160 165 170 175	585
gtc aaa tgg acc gag gcg acc gcg atc ctt gcg ggc gat gcg ctg cag Val Lys Trp Thr Glu Ala Thr Ala Ile Leu Ala Gly Asp Ala Leu Gln 180 185 190	633
acg ctg gcc ttc cag ctg ctg gcc gat ccg cgc gtg ggc gac gat gcg Thr Leu Ala Phe Gln Leu Leu Ala Asp Pro Arg Val Gly Asp Asp Ala 195 200 205	681
gcf cgg atg cgg ctg ggt tcg ctg gcg cag gca tcg ggg gct gcf Ala Arg Met Arg Leu Val Gly Ser Leu Ala Gln Ala Ser Gly Ala Ala 210 215 220	729
ggc atg gtc tgg ggc cag gcg ctg gac atc gcg gcc gag acc tcg ggc Gly Met Val Trp Gly Gln Ala Leu Asp Ile Ala Ala Glu Thr Ser Gly 225 230 235	777
tgf ccc ctg gat ctg gac gcg atc atc cgc ctg cag ggt ggc aag acc Val Pro Leu Asp Leu Asp Ala Ile Ile Arg Leu Gln Gly Gly Lys Thr 240 245 250 255	825
ggc gcg ctg atc cgc ttt gcc gcg acc gcc ggg ccc ctg atg gcg ggg Gly Ala Leu Ile Arg Phe Ala Ala Thr Ala Gly Pro Leu Met Ala Gly 260 265 270	873
gcf gac cct gcc gcg ctg gac gat tat gcf cag gcc gtc ggg ctg gcc Ala Asp Pro Ala Ala Leu Asp Asp Tyr Ala Gln Ala Val Gly Leu Ala 275 280 285	921
ttc cag atc gcf gac gac atc ctg gac gtc gag ggc tgc gag gcc gcf Phe Gln Ile Ala Asp Asp Ile Leu Asp Val Glu Gly Cys Glu Ala Ala 290 295 300	969
acc ggc aag cgc gtc ggc aag gat gcf gat gcc aac aag gcf acc ttc Thr Gly Lys Arg Val Gly Lys Asp Ala Asp Ala Asn Lys Ala Thr Phe 305 310 315	1017
gtc tcg ctg ctg ggc ctc gag ggg gcf cgg tcc gag gcf cgt cgc ctg	1065

Val Ser Leu Leu Gly Leu Glu Gly Ala Arg Ser Glu Ala Arg Arg Leu				
320	325	330	335	
gcc gat gcg ggg cag gac gcg ctg gcg ggt tac ggc gat gct gct gcg ggg				1113
Ala Asp Ala Gly Gln Asp Ala Leu Ala Gly Tyr Gly Asp Ala Ala Gly				
340	345	350		
aac ctt cg gac ctg gcg cgc ttc gtg atc gaa cgc gac agc tga				1158
Asn Leu Arg Asp Leu Ala Arg Phe Val Ile Glu Arg Asp Ser				
355	360	365		
tcgcccctt cccgccaagg ggcaag atg atg acc gac gga ccc gca acc ccg				1211
Met Met Thr Asp Gly Pro Ala Thr Pro				
370				
atc ctg gac cgc gtc cag cag cca tcc gac ctg gca tcg ctg gac gat				1259
Ile Leu Asp Arg Val Gln Gln Pro Ser Asp Leu Ala Ser Leu Asp Asp				
375	380	385	390	
gcg cag ctg cgc ctg ctg gcg gac gag ctg cgg gcc gag acc atc gac				1307
Ala Gln Leu Arg Leu Leu Ala Asp Glu Leu Arg Ala Glu Thr Ile Asp				
395	400	405		
atc gtc acg cgc acg ggc ggt cac ctg ggc gcg ggg ctg ggc gtg gtc				1355
Ile Val Ser Arg Thr Gly Gly His Leu Gly Ala Gly Leu Gly Val Val				
410	415	420		
gaa ctg acg gtc gcc ctg cac gcc gtc ttt cgg gcg ccg cgc gac aag				1403
Glu Leu Thr Val Ala Leu His Ala Val Phe Arg Ala Pro Arg Asp Lys				
425	430	435		
atc gtc tgg gac gtg ggg cat caa tgc tat ccc cac aag atc ctg acg				1451
Ile Val Trp Asp Val Gly His Gln Cys Tyr Pro His Lys Ile Leu Thr				
440	445	450		
ggc agg cgg gac cgg atg cgc acg ctg cgc atg ggc ggc ggg ctg tcg				1499
Gly Arg Arg Asp Arg Met Arg Thr Leu Arg Met Gly Gly Gly Leu Ser				
455	460	465	470	
ggg ttc acc aag cgg cag gaa agc gcg ttc gat ccg ttc ggt gcg ggg				1547
Gly Phe Thr Lys Arg Gln Glu Ser Ala Phe Asp Pro Phe Gly Ala Gly				
475	480	485		
cac agc tcg acc tcg atc tcg gcg gcg ctg ggc ttc gcg atg gcg cgt				1595
His Ser Ser Thr Ser Ile Ser Ala Ala Leu Gly Phe Ala Met Ala Arg				
490	495	500		
gaa ctt ggc ggg gat cc				1612
Glu Leu Gly Gly Asp				
505				

<210> 158

<211> 78

<212> PRT

<213> Paracoccus sp. R114

<400> 158

Met Ser Asp Ile Gln Thr Leu Ser Phe Glu Glu Ala Met Arg Glu Leu
5 10 15
1

Glu Ala Thr Val Gly Lys Leu Glu Thr Gly Glu Ala Thr Leu Glu Asp
20 25 30

Ser Ile Ala Leu Tyr Glu Arg Gly Ala Ala Leu Arg Ala His Cys Glu
35 40 45

Thr Arg Leu Arg Glu Ala Glu Glu Arg Val Glu Lys Ile Thr Leu Ala
50 55 60

Ala Asn Gly Gln Pro Ser Gly Thr Glu Pro Ala Glu Gly Leu
65 70 75

<210> 159

<211> 287

<212> PRT

<213> Paracoccus sp. R114

<400> 159

Met Gln Ala Arg Leu Ala Glu Ile Arg Pro Leu Val Glu Ala Glu Leu
5 10 15
1

Asn Ala Ala Ile Asp Ala Leu Pro Ala Gly Asp Leu Ser Asp Ala Met
20 25 30

Arg Tyr Ala Val Gln Gly Gly Lys Arg Leu Arg Ala Phe Leu Val Met
35 40 45

Glu Ser Ala Arg Leu His Gly Leu Asp Asp Asp Ala Ser Leu Pro Val
50 55 60

Ala Ala Ala Val Glu Ala Leu His Ala Tyr Ser Leu Val His Asp Asp
65 70 75 80

Leu Pro Ala Met Asp Asp Asp Asp Leu Arg Arg Gly Gln Pro Thr Val
85 90 95

His Val Lys Trp Thr Glu Ala Thr Ala Ile Leu Ala Gly Asp Ala Leu
100 105 110

Gln Thr Leu Ala Phe Gln Leu Leu Ala Asp Pro Arg Val Gly Asp Asp
115 120 125

Ala Ala Arg Met Arg Leu Val Gly Ser Leu Ala Gln Ala Ser Gly Ala
130 135 140

Ala Gly Met Val Trp Gly Gln Ala Leu Asp Ile Ala Ala Glu Thr Ser
145 150 155 160

Gly Val Pro Leu Asp Leu Asp Ala Ile Ile Arg Leu Gln Gly Gly Lys
165 170 175

Thr Gly Ala Leu Ile Arg Phe Ala Ala Thr Ala Gly Pro Leu Met Ala
180 185 190

Gly Ala Asp Pro Ala Ala Leu Asp Asp Tyr Ala Gln Ala Val Gly Leu
195 200 205

Ala Phe Gln Ile Ala Asp Asp Ile Leu Asp Val Glu Gly Cys Glu Ala
210 215 220

Ala Thr Gly Lys Arg Val Gly Lys Asp Ala Asp Ala Asn Lys Ala Thr
225 230 235 240

Phe Val Ser Leu Leu Gly Leu Glu Gly Ala Arg Ser Glu Ala Arg Arg
245 250 255

Leu Ala Asp Ala Gly Gln Asp Ala Leu Ala Gly Tyr Gly Asp Ala Ala
260 265 270

Gly Asn Leu Arg Asp Leu Ala Arg Phe Val Ile Glu Arg Asp Ser
275 280 285

<210> 160

<211> 142

<212> PRT

<213> Paracoccus sp. R114

<400> 160

Met Met Thr Asp Gly Pro Ala Thr Pro Ile Leu Asp Arg Val Gln Gln
1 5 10 15

Pro Ser Asp Leu Ala Ser Leu Asp Asp Ala Gln Leu Arg Leu Leu Ala
20 25 30

Asp Glu Leu Arg Ala Glu Thr Ile Asp Ile Val Ser Arg Thr Gly Gly
35 40 45

His Leu Gly Ala Gly Leu Gly Val Val Glu Leu Thr Val Ala Leu His
50 55 60

Ala Val Phe Arg Ala Pro Arg Asp Lys Ile Val Trp Asp Val Gly His
65 70 75 80

Gln Cys Tyr Pro His Lys Ile Leu Thr Gly Arg Arg Asp Arg Met Arg
85 90 95

Thr Leu Arg Met Gly Gly Leu Ser Gly Phe Thr Lys Arg Gln Glu
100 105 110

Ser Ala Phe Asp Pro Phe Gly Ala Gly His Ser Ser Thr Ser Ile Ser
115 120 125

Ala Ala Leu Gly Phe Ala Met Ala Arg Glu Leu Gly Gly Asp
130 135 140

<210> 161

<211> 6

<212> PRT

<213> Bradyrhizobium japonicum

<400> 161

Val His Asp Asp Leu Pro
1 5

<210> 162

<211> 6

<212> PRT

<213> Rhizobium sp. strain NGR234

<400> 162

Val His Asp Asp Leu Pro
1 5

<210> 163

<211> 6

<212> PRT

<213> Bacillus stearothermophilus

<400> 163

Ile His Asp Asp Leu Pro
1 5

<210> 164

<211> 6

<212> PRT

<213> Bacillus subtilis

<400> 164

Ile His Asp Asp Leu Pro
1 5

<210> 165

<211> 6

<212> PRT

<213> Escherichia coli

<400> 165

Ile His Asp Asp Leu Pro
1 5

<210> 166

<211> 6

<212> PRT

<213> Haemophilus influenzae

<400> 166

Ile His Asp Asp Leu Pro
1 5

<210> 167

<211> 16

<212> DNA

<213> synthetic construct

<400> 167

tccaygayga yctgcc

16

<210> 168

<211> 5

<212> PRT

<213> Bradyrhizobium japonicum

<400> 168

Asp Asp Ile Leu Asp
1 5

<210> 169

<211> 5

<212> PRT

<213> Rhizobium sp. strain NGR234

<400> 169

Asp Asp Ile Leu Asp
1 5

<210> 170

<211> 5

<212> PRT

<213> *Bacillus stearothermophilus*

C <400> 170

Asp Asp Ile Leu Asp
1 5

<210> 171

<211> 5

<212> PRT

<213> *Bacillus subtilis*

<400> 171

Asp Asp Ile Leu Asp
1 5

C <210> 172

<211> 5

<212> PRT

<213> *Escherichia coli*

<400> 172

Asp Asp Ile Leu Asp
1 5

<210> 173

<211> 5

<212> PRT

<213> Haemophilus influenzae

<400> 173

Asp Asp Ile Leu Asp
1 5

<210> 174

<211> 15

<212> DNA

<213> synthetic construct

<400> 174
gaygayatcc tggay

15

<210> 175

<211> 1176

<212> DNA

<213> Paracoccus sp. R1534

<220>

<221> CDS

<222> (1)..(1173)

<223> acetyl-CoA acetyltransferase

<400> 175
atg gac ccc atc gtc atc acc ggc gcg atg cgc acc ccg atg ggg gca 48
Met Asp Pro Ile Val Ile Thr Gly Ala Met Arg Thr Pro Met Gly Ala
5 10 15
ttc cag ggc gat ctt gcc gcg atg gat gcc ccg acc ctt ggc gcg gcc 96
Phe Gln Gly Asp Leu Ala Ala Met Asp Ala Pro Thr Leu Gly Ala Ala
20 25 30
gcg atc cgc gcc gcg ctg aac ggc ctg ccc gac atg gtg gac gag 144
Ala Ile Arg Ala Ala Leu Asn Gly Leu Ser Pro Asp Met Val Asp Glu
35 40 45

gtg ctg atg ggc tgc gtc ctg ccc gcg ggc cag ggt cag gca ccg gca Val Leu Met Gly Cys Val Leu Pro Ala Gly Gln Gly Gln Ala Pro Ala	192
50 55 60	
cgt cag gcg gcg ctt gac gcc gga ctg ccg ctg tcg gcg ggc gcg acc Arg Gln Ala Ala Leu Asp Ala Gly Leu Pro Leu Ser Ala Gly Ala Thr	240
65 70 75 80	
acc atc aac aag atg tgc gga tcg ggc atg aag gcc gcg atg ctg ggc Thr Ile Asn Lys Met Cys Gly Ser Gly Met Lys Ala Ala Met Leu Gly	288
85 90 95	
cat gac ctg atc gcc gcg gga tcg gcg ggc atc gtc gtc gcc ggc ggg His Asp Leu Ile Ala Ala Gly Ser Ala Gly Ile Val Val Ala Gly Gly	336
100 105 110	
atg gag agc atg tcg aac gcc ccc tac ctg ctg ccc aag gcg cgg tcg Met Glu Ser Met Ser Asn Ala Pro Tyr Leu Leu Pro Lys Ala Arg Ser	384
115 120 125	
ggg atg cgc atg ggc cat gac cgt gtg ctg gat cac atg ttc ctc gac Gly Met Arg Met Gly His Asp Arg Val Leu Asp His Met Phe Leu Asp	432
130 135 140	
ggg ttg gag gac gcc tat gac aag ggc cgc ctg atg ggc acc ttc gcc Gly Leu Glu Asp Ala Tyr Asp Lys Gly Arg Leu Met Gly Thr Phe Ala	480
145 150 155 160	
gag gat tgc gcc ggc gat cac ggt ttc acc cgc gag gcg cag gac gac Glu Asp Cys Ala Gly Asp His Gly Phe Thr Arg Glu Ala Gln Asp Asp	528
165 170 175	
tat gcg ctg acc agc ctg gcc cgc gcg cag gac gcc atc gcc agc ggt Tyr Ala Leu Thr Ser Leu Ala Arg Ala Gln Asp Ala Ile Ala Ser Gly	576
180 185 190	
gcc ttc gcc gcc gag atc gcg ccc gtg acc gtc acg gca cgc aag gtg Ala Phe Ala Ala Glu Ile Ala Pro Val Thr Val Thr Ala Arg Lys Val	624
195 200 205	
cag acc acc gtc gat acc gac gag atg ccc ggc aag gcc cgc ccc gag Gln Thr Thr Val Asp Thr Asp Glu Met Pro Gly Lys Ala Arg Pro Glu	672
210 215 220	
aag atc ccc cat ctg aag ccc gcc ttc cgt gac ggt ggc acg gtc acg Lys Ile Pro His Leu Lys Pro Ala Phe Arg Asp Gly Gly Thr Val Thr	720
225 230 235 240	
gcg gcg aac agc tcg atc tcg gac ggg ggc gcg gcg ctg gtg atg Ala Ala Asn Ser Ser Ile Ser Asp Gly Ala Ala Leu Val Met	768
245 250 255	
atg cgc cag tcg cag gcc gag aag ctg ggc ctg acg ccg atc gcg cgg Met Arg Gln Ser Gln Ala Glu Lys Leu Gly Leu Thr Pro Ile Ala Arg	816
260 265 270	

atc atc ggt cat gcg acc cat gcc gac cgt ccc ggc ctg ttc ccg acg Ile Ile Gly His Ala Thr His Ala Asp Arg Pro Gly Leu Phe Pro Thr 275 280 285	864
gcc ccc atc ggc gcg atg cgc aag ctg ctg gac cgc acg gac acc cgc Ala Pro Ile Gly Ala Met Arg Lys Leu Leu Asp Arg Thr Asp Thr Arg 290 295 300	912
ctt ggc gat tac gac ctg ttc gag gtg aac gag gca ttc gcc gtc gtc Leu Gly Asp Tyr Asp Leu Phe Glu Val Asn Glu Ala Phe Ala Val Val 305 310 315 320	960
gcc atg atc gcg atg aag gag ctt ggc ctg cca cac gat gcc acg aac Ala Met Ile Ala Met Lys Glu Leu Gly Leu Pro His Asp Ala Thr Asn 325 330 335	1008
atc aac ggc ggg gcc tgc gcg ctt ggg cat ccc atc ggc gcg tcg ggg Ile Asn Gly Gly Ala Cys Ala Leu Gly His Pro Ile Gly Ala Ser Gly 340 345 350	1056
gcg cgg atc atg gtc acg ctg ctg aac gcg atg gcg gcg cgg ggc gcg Ala Arg Ile Met Val Thr Leu Leu Asn Ala Met Ala Ala Arg Gly Ala 355 360 365	1104
acg cgc ggg gcc gca tcc gtc tgc atc ggc ggg ggc gag gcg acg gcc Thr Arg Gly Ala Ala Ser Val Cys Ile Gly Gly Glu Ala Thr Ala 370 375 380	1152
atc gcg ctg gaa cgg ctg agc taa Ile Ala Leu Glu Arg Leu Ser 385 390	1176
<210> 176	
<211> 391	
<212> PRT	
<213> Paracoccus sp. R1534	
<400> 176	
Met Asp Pro Ile Val Ile Thr Gly Ala Met Arg Thr Pro Met Gly Ala 1 5 10 15	
Phe Gln Gly Asp Leu Ala Ala Met Asp Ala Pro Thr Leu Gly Ala Ala 20 25 30	
Ala Ile Arg Ala Ala Leu Asn Gly Leu Ser Pro Asp Met Val Asp Glu 35 40 45	

Val Leu Met Gly Cys Val Leu Pro Ala Gly Gln Gly Gln Ala Pro Ala
50 55 60

Arg Gln Ala Ala Leu Asp Ala Gly Leu Pro Leu Ser Ala Gly Ala Thr
65 70 75 80

Thr Ile Asn Lys Met Cys Gly Ser Gly Met Lys Ala Ala Met Leu Gly
85 90 95

His Asp Leu Ile Ala Ala Gly Ser Ala Gly Ile Val Val Ala Gly Gly
100 105 110

Met Glu Ser Met Ser Asn Ala Pro Tyr Leu Leu Pro Lys Ala Arg Ser
115 120 125

Gly Met Arg Met Gly His Asp Arg Val Leu Asp His Met Phe Leu Asp
130 135 140

Gly Leu Glu Asp Ala Tyr Asp Lys Gly Arg Leu Met Gly Thr Phe Ala
145 150 155 160

Glu Asp Cys Ala Gly Asp His Gly Phe Thr Arg Glu Ala Gln Asp Asp
165 170 175

Tyr Ala Leu Thr Ser Leu Ala Arg Ala Gln Asp Ala Ile Ala Ser Gly
180 185 190

Ala Phe Ala Ala Glu Ile Ala Pro Val Thr Val Thr Ala Arg Lys Val
195 200 205

Gln Thr Thr Val Asp Thr Asp Glu Met Pro Gly Lys Ala Arg Pro Glu
210 215 220

Lys Ile Pro His Leu Lys Pro Ala Phe Arg Asp Gly Gly Thr Val Thr
225 230 235 240

Ala Ala Asn Ser Ser Ile Ser Asp Gly Ala Ala Ala Leu Val Met
245 250 255

Met Arg Gln Ser Gln Ala Glu Lys Leu Gly Leu Thr Pro Ile Ala Arg
260 265 270

Ile Ile Gly His Ala Thr His Ala Asp Arg Pro Gly Leu Phe Pro Thr

275

280

285

Ala Pro Ile Gly Ala Met Arg Lys Leu Leu Asp Arg Thr Asp Thr Arg
290 295 300

Leu Gly Asp Tyr Asp Leu Phe Glu Val Asn Glu Ala Phe Ala Val Val
305 310 315 320

Ala Met Ile Ala Met Lys Glu Leu Gly Leu Pro His Asp Ala Thr Asn
325 330 335

Ile Asn Gly Gly Ala Cys Ala Leu Gly His Pro Ile Gly Ala Ser Gly
340 345 350

Ala Arg Ile Met Val Thr Leu Leu Asn Ala Met Ala Ala Arg Gly Ala
355 360 365

Thr Arg Gly Ala Ala Ser Val Cys Ile Gly Gly Glu Ala Thr Ala
370 375 380

Ile Ala Leu Glu Arg Leu Ser
385 390

<210> 177

<211> 1980

<212> DNA

<213> Paracoccus sp. R114

<220>

<221> CDS

<222> (1)..(1170)

<223> phaA

<220>

<221> misc_feature

<222> (1179)..(1194)

<223> inverted repeat between genes constituting a putative transcripti
onal sto

<220>

<221> misc_feature

<222> (1196)..(1210)

<223> inverted repeat between genes constituting a putative transcripti
onal sto

<220>

<221> CDS

<222> (1258)..(1980)

<223> phaB

<400> 177

atg acc aaa gcc gta atc gta tct gcc gca cgt acc ccc gtc ggc agc	48
Met Thr Lys Ala Val Ile Val Ser Ala Ala Arg Thr Pro Val Gly Ser	
1 5 10 15	

ttc atg ggc gca ttc gcc aat gtc ccc gca cat gat ctg ggc gcc gcc	96
Phe Met Gly Ala Phe Ala Asn Val Pro Ala His Asp Leu Gly Ala Ala	
20 25 30	

gtc ctg cgc gag gtc gtg gcc cgc gcc ggt gtc gac ccc gcc gag gtc	144
Val Leu Arg Glu Val Val Ala Arg Ala Gly Val Asp Pro Ala Glu Val	
35 40 45	

agc gag acg atc ctg ggc cag gtg ctg acc gcc gcg cag ggc cag aac	192
Ser Glu Thr Ile Leu Gly Gln Val Leu Thr Ala Ala Gln Gly Gln Asn	
50 55 60	

ccc gcg cgc cag gcg cat atc aat gcg ggc ctg ccc aag gaa tcg gcg	240
Pro Ala Arg Gln Ala His Ile Asn Ala Gly Leu Pro Lys Glu Ser Ala	
65 70 75 80	

gcg tgg ctc atc aac cag gtc tgc ggc tcg ggg ctg cgc gcc gtc gcg	288
Ala Trp Leu Ile Asn Gln Val Cys Gly Ser Gly Leu Arg Ala Val Ala	
85 90 95	

ctg gcg cgc cag cag gtc atg ctg ggc gat gcg cag atc gtt ctg gcg	336
Leu Ala Ala Gln Gln Val Met Leu Gly Asp Ala Gln Ile Val Leu Ala	
100 105 110	

ggg ggc cag gag agc atg tcg ctg tcg acc cat gcc gcc tat ctg cgc	384
Gly Gly Gln Glu Ser Met Ser Leu Ser Thr His Ala Ala Tyr Leu Arg	

115	120	125	
gct ggc cag aag atg ggc gac atg aag atg atc gac acc atc cgc Ala Gly Gln Lys Met Gly Asp Met Lys Met Ile Asp Thr Met Ile Arg 130 135 140			
gac ggg ctg tgg gat gcc ttc aac ggc tat cac atg ggt cag acc gcc Asp Gly Leu Trp Asp Ala Phe Asn Gly Tyr His Met Gly Gln Thr Ala 145 150 155 160			
gag aac gtg gcc gac cag tgg tcg atc agc cgc gac cag cag gac gaa Glu Asn Val Ala Asp Gln Trp Ser Ile Ser Arg Asp Gln Gln Asp Glu 165 170 175			
ttc gcc ctg gct tcg cag aac aag gcc gag gcc gcg cag aat gct ggc Phe Ala Leu Ala Ser Gln Asn Lys Ala Glu Ala Ala Gln Asn Ala Gly 180 185 190			
cgc ttc gat gac gaa atc gtc gcc tat acc gtc aag ggc cgc aag ggc Arg Phe Asp Asp Glu Ile Val Ala Tyr Thr Val Lys Gly Arg Lys Gly 195 200 205			
gac acg gtc gtc gac aag gac gaa tac atc cgc cac ggc gcc acg atc Asp Thr Val Val Asp Lys Asp Glu Tyr Ile Arg His Gly Ala Thr Ile 210 215 220			
gag ggc atg cag aag ctg cgc ccc gcc ttc acc aag gaa ggc tcg gtc Glu Gly Met Gln Lys Leu Arg Pro Ala Phe Thr Lys Glu Gly Ser Val 225 230 235 240			
acg gcg ggc aac gcg tcg ggc ctg aac gac ggc gcg gcg gcc gtc atg Thr Ala Gly Asn Ala Ser Gly Leu Asn Asp Gly Ala Ala Ala Val Met 245 250 255			
gtc atg tcc gag gac gag gcc gca cgc cgc ggg ctg acg ccg ctg gcg Val Met Ser Glu Asp Glu Ala Ala Arg Arg Gly Leu Thr Pro Leu Ala 260 265 270			
cgc atc gcc tcc tat gcg acg gcg ggc ctc gac ccg gcg atc atg ggc Arg Ile Ala Ser Tyr Ala Thr Ala Gly Leu Asp Pro Ala Ile Met Gly 275 280 285			
acc ggg ccg atc ccc tcc agc cgc aag gcg ctg gaa aag gcg ggc tgg Thr Gly Pro Ile Pro Ser Ser Arg Lys Ala Leu Glu Lys Ala Gly Trp 290 295 300			
tcg gtc ggc gac ctg gac ctg gag gcg aac gag gcc ttt gcc gcg Ser Val Gly Asp Leu Asp Leu Val Glu Ala Asn Glu Ala Phe Ala Ala 305 310 315 320			
cag gcc tgc gcc gtg aac aag gac atg ggc tgg gat ccg tcc atc gtg Gln Ala Cys Ala Val Asn Lys Asp Met Gly Trp Asp Pro Ser Ile Val 325 330 335			
aac gtc aac ggc ggc gcg atc gcc atc ggc cac ccg atc ggc gcc tcg Asn Val Asn Gly Gly Ala Ile Ala Ile Gly His Pro Ile Gly Ala Ser 340 345 350			

ggg gcg cgg atc ctg aac acc ctg ctg ttc gag atg cag cgc cgc gac Gly Ala Arg Ile Leu Asn Thr Leu Leu Phe Glu Met Gln Arg Arg Asp 355	360	365	1104
gcc aag aag ggc ctt gcg acg ctg tgc atc ggc ggc ggc atg ggc gtc Ala Lys Lys Gly Leu Ala Thr Leu Cys Ile Gly Gly Gly Met Gly Val 370	375	380	1152
gcc atg tgc ctc gaa cgc tgaacgaccg gcgtgtgcgc aatttaattg Ala Met Cys Leu Glu Arg 385	390		1200
cgcacacgcc ccctgcaaag tagcaatgtt ttacgataac gaatgaaggg gggaaatc			1257
atg tcc aag gta gca ctg gtc acc ggc gga tcg cgc ggc atc ggc gcc Met Ser Lys Val Ala Leu Val Thr Gly Gly Ser Arg Gly Ile Gly Ala 395	400	405	1305
gag atc tgc aag gcg ctt cag gcc gca ggc tat acc gtc gcc gcg aac Glu Ile Cys Lys Ala Leu Gln Ala Ala Gly Tyr Thr Val Ala Ala Asn 410	415	420	1353
tat gcc ggc aat gac gac gcg gcc aag gcc ttc acc gag gaa acc acc ggc Tyr Ala Gly Asn Asp Asp Ala Ala Lys Ala Phe Thr Glu Glu Thr Gly 425	430	435	1401
atc aag acc tac aag tgg tcg gtc gcc gat tac gat gcc tgc aag gcc Ile Lys Thr Tyr Lys Trp Ser Val Ala Asp Tyr Asp Ala Cys Lys Ala 440	445	450	1449
ggc atc gcc cag gtc gaa gag gat ctg ggc ccc atc gcc gtg ctg atc Gly Ile Ala Gln Val Glu Glu Asp Leu Gly Pro Ile Ala Val Leu Ile 455	460	465	1497
aac aat gcc ggg atc acc cgc gac gcg ccc ttc cac aag atg acg ccc Asn Asn Ala Gly Ile Thr Arg Asp Ala Pro Phe His Lys Met Thr Pro 475	480	485	1545
gag aag tgg aag gag gtc atc gac acc aac ctg acc ggc acc ttc aac Glu Lys Trp Lys Glu Val Ile Asp Thr Asn Leu Thr Gly Thr Phe Asn 490	495	500	1593
atg acc cat ccg gtc tgg ccc ggc atg cgc gaa cgc aag ttc gga cgc Met Thr His Pro Val Trp Pro Gly Met Arg Glu Arg Lys Phe Gly Arg 505	510	515	1641
gtc atc aac atc agc tcg atc aac ggg cag aag ggc cag ttc ggg cag Val Ile Asn Ile Ser Ser Ile Asn Gly Gln Lys Gly Gln Phe Gly Gln 520	525	530	1689
gcg aac tat gcc gcg gcc aag gcg ggc gac ctg ggc ttc acc aag tcg Ala Asn Tyr Ala Ala Lys Ala Gly Asp Leu Gly Phe Thr Lys Ser 535	540	545	1737
ctg gcg cag gaa ggc gcg cgc aac aac atc acc gtc aac gcg atc tgc Leu Ala Gln Glu Gly Ala Arg Asn Asn Ile Thr Val Asn Ala Ile Cys			1785

	555	560	565	
ccc ggc tat atc gcg acg gac atg gtg atg gcc gtt ccc gaa cag gtc Pro Gly Tyr Ile Ala Thr Asp Met Val Met Ala Val Pro Glu Gln Val	570	575	580	1833
cgc gag ggg atc atc gcg cag atc ccc gtc ggc cgcc ttg ggc gag ccc Arg Glu Gly Ile Ile Ala Gln Ile Pro Val Gly Arg Leu Gly Glu Pro	585	590	595	1881
tcc gag atc gcg cgc tgc gtg gtg ttc ctg gcc tcc gac gat gcg ggc Ser Glu Ile Ala Arg Cys Val Val Phe Leu Ala Ser Asp Asp Ala Gly	600	605	610	1929
ttc gtc aca ggc tcg acc atc acg gcg aat ggc ggc cag tac tac atc Phe Val Thr Gly Ser Thr Ile Thr Ala Asn Gly Gly Gln Tyr Tyr Ile	615	620	625	1977
630				
tga				1980

<210> 178

<211> 390

<212> PRT

<213> Paracoccus sp. R114

<220>

<221> misc_feature

<222> (1179)...(1194)

<223> inverted repeat between genes constituting a putative transcripti
onal sto

<220>

<221> misc_feature

<222> (1196)...(1210)

<223> inverted repeat between genes constituting a putative transcripti
onal sto

<400> 178

Met Thr Lys Ala Val Ile Val Ser Ala Ala Arg Thr Pro Val Gly Ser			
1	5	10	15

Phe Met Gly Ala Phe Ala Asn Val Pro Ala His Asp Leu Gly Ala Ala			
20	25	30	

Val Leu Arg Glu Val Val Ala Arg Ala Gly Val Asp Pro Ala Glu Val
35 40 45

Ser Glu Thr Ile Leu Gly Gln Val Leu Thr Ala Ala Gln Gly Gln Asn
50 55 60

Pro Ala Arg Gln Ala His Ile Asn Ala Gly Leu Pro Lys Glu Ser Ala
65 70 75 80

Ala Trp Leu Ile Asn Gln Val Cys Gly Ser Gly Leu Arg Ala Val Ala
85 90 95

Leu Ala Ala Gln Gln Val Met Leu Gly Asp Ala Gln Ile Val Leu Ala
100 105 110

Gly Gly Gln Glu Ser Met Ser Leu Ser Thr His Ala Ala Tyr Leu Arg
115 120 125

Ala Gly Gln Lys Met Gly Asp Met Lys Met Ile Asp Thr Met Ile Arg
130 135 140

Asp Gly Leu Trp Asp Ala Phe Asn Gly Tyr His Met Gly Gln Thr Ala
145 150 155 160

Glu Asn Val Ala Asp Gln Trp Ser Ile Ser Arg Asp Gln Gln Asp Glu
165 170 175

Phe Ala Leu Ala Ser Gln Asn Lys Ala Glu Ala Ala Gln Asn Ala Gly
180 185 190

Arg Phe Asp Asp Glu Ile Val Ala Tyr Thr Val Lys Gly Arg Lys Gly
195 200 205

Asp Thr Val Val Asp Lys Asp Glu Tyr Ile Arg His Gly Ala Thr Ile
210 215 220

Glu Gly Met Gln Lys Leu Arg Pro Ala Phe Thr Lys Glu Gly Ser Val
225 230 235 240

Thr Ala Gly Asn Ala Ser Gly Leu Asn Asp Gly Ala Ala Ala Val Met
245 250 255

Val Met Ser Glu Asp Glu Ala Ala Arg Arg Gly Leu Thr Pro Leu Ala
260 265 270

Arg Ile Ala Ser Tyr Ala Thr Ala Gly Leu Asp Pro Ala Ile Met Gly
275 280 285

Thr Gly Pro Ile Pro Ser Ser Arg Lys Ala Leu Glu Lys Ala Gly Trp
290 295 300

Ser Val Gly Asp Leu Asp Leu Val Glu Ala Asn Glu Ala Phe Ala Ala
305 310 315 320

Gln Ala Cys Ala Val Asn Lys Asp Met Gly Trp Asp Pro Ser Ile Val
325 330 335

Asn Val Asn Gly Gly Ala Ile Ala Ile Gly His Pro Ile Gly Ala Ser
340 345 350

Gly Ala Arg Ile Leu Asn Thr Leu Leu Phe Glu Met Gln Arg Arg Asp
355 360 365

Ala Lys Lys Gly Leu Ala Thr Leu Cys Ile Gly Gly Met Gly Val
370 375 380

Ala Met Cys Leu Glu Arg
385 390

<210> 179

<211> 240

<212> PRT

<213> Paracoccus sp. R114

<220>

<221> misc_feature

<222> (1179)..(1194)

<223> inverted repeat between genes constituting a putative transcripti
onal sto

<220>

<221> misc_feature

<222> (1196)..(1210)

<223> inverted repeat between genes constituting a putative transcripti
onal sto

<400> 179

Met Ser Lys Val Ala Leu Val Thr Gly Gly Ser Arg Gly Ile Gly Ala
1 5 10 15

Glu Ile Cys Lys Ala Leu Gln Ala Ala Gly Tyr Thr Val Ala Ala Asn
20 25 30

Tyr Ala Gly Asn Asp Asp Ala Ala Lys Ala Phe Thr Glu Glu Thr Gly
35 40 45

Ile Lys Thr Tyr Lys Trp Ser Val Ala Asp Tyr Asp Ala Cys Lys Ala
50 55 60

Gly Ile Ala Gln Val Glu Glu Asp Leu Gly Pro Ile Ala Val Leu Ile
65 70 75 80

Asn Asn Ala Gly Ile Thr Arg Asp Ala Pro Phe His Lys Met Thr Pro
85 90 95

Glu Lys Trp Lys Glu Val Ile Asp Thr Asn Leu Thr Gly Thr Phe Asn
100 105 110

Met Thr His Pro Val Trp Pro Gly Met Arg Glu Arg Lys Phe Gly Arg
115 120 125

Val Ile Asn Ile Ser Ser Ile Asn Gly Gln Lys Gly Gln Phe Gly Gln
130 135 140

Ala Asn Tyr Ala Ala Lys Ala Gly Asp Leu Gly Phe Thr Lys Ser
145 150 155 160

Leu Ala Gln Glu Gly Ala Arg Asn Asn Ile Thr Val Asn Ala Ile Cys
165 170 175

Pro Gly Tyr Ile Ala Thr Asp Met Val Met Ala Val Pro Glu Gln Val
180 185 190

Arg Glu Gly Ile Ile Ala Gln Ile Pro Val Gly Arg Leu Gly Glu Pro
195 200 205

Ser Glu Ile Ala Arg Cys Val Val Phe Leu Ala Ser Asp Asp Ala Gly
210 215 220

Phe Val Thr Gly Ser Thr Ile Thr Ala Asn Gly Gly Gln Tyr Tyr Ile
225 230 235 240

<210> 180

<211> 729

<212> DNA

<213> Paracoccus carotinifaciens E-396

<220>

<221> CDS

<222> (1) .. (726)

<223> Beta-carotene Beta-4 oxygenase

atc gtc tcg ggc ggc atc atc gcc gcg tgg ctg gcc ctg cat gtg cat
 Ile Val Ser Gly Gly Ile Ile Ala Ala Trp Leu Ala Leu His Val His
 20 25 30

gcg ctg tgg ttt ctg gac gcg gcg gcg cat ccc atc ctg gcg gtc gcg 144
 Ala Leu Trp Phe Leu Asp Ala Ala Ala His Pro Ile Leu Ala Val Ala
 35 40 45

```

aat ttc ctg ggg ctg acc tgg ctg tcg gtc ggt ctg ttc atc atc gcg      192
Asn Phe Leu Gly Leu Thr Trp Leu Ser Val Gly Leu Phe Ile Ile Ala
50          55          60

```

```

cat gac gcg atg cat ggg tcg gtc gtg ccg ggg cgc ccg cgc gcc aat      240
His Asp Ala Met His Gly Ser Val Val Pro Gly Arg Pro Arg Ala Asn
65                      70                         75                     80

```

gct gct atg ggc cag ctt gtc ctg tgg ctg tat gcc gga ttt tcc tgg
 Ala Ala Met Gly Gln Leu Val Leu Trp Leu Tyr Ala Gly Phe Ser Trp
 85 90 95

cgc aag atg atc gtc aag cac atg gcc cat cat cgc cat gcc gga acc Arg Lys Met Ile Val Lys His Met Ala His His Arg His Ala Gly Thr 100 105 110	336
gac gac gac cca gat ttc gac cat ggc ggc ccg gtc cgc tgg tac gcc Asp Asp Asp Pro Asp Phe Asp His Gly Gly Pro Val Arg Trp Tyr Ala 115 120 125	384
cgc ttc atc ggc acc tat ttc ggc tgg cgc gag ggg ctg ctg ctg ccc Arg Phe Ile Gly Thr Tyr Phe Gly Trp Arg Glu Gly Leu Leu Leu Pro 130 135 140	432
gtc atc gtg acg gtc tat gcg ctg atg ttg ggg gat cgc tgg atg tac Val Ile Val Thr Val Tyr Ala Leu Met Leu Gly Asp Arg Trp Met Tyr 145 150 155 160	480
gtg gtc ttc tgg ccg ttg ccg tcg atc ctg gcg tcg atc cag ctg ttc Val Val Phe Trp Pro Leu Pro Ser Ile Leu Ala Ser Ile Gln Leu Phe 165 170 175	528
gtg ttc ggc atc tgg ctg ccg cac cgc ccc ggc cac gac gcg ttc ccg Val Phe Gly Ile Trp Leu Pro His Arg Pro Gly His Asp Ala Phe Pro 180 185 190	576
gac cgc cac aat gcg cgg tcg tcg cgg atc agc gac ccc gtg tcg ctg Asp Arg His Asn Ala Arg Ser Ser Arg Ile Ser Asp Pro Val Ser Leu 195 200 205	624
ctg acc tgc ttt cac ttt ggc ggt tat cat cac gaa cac cac ctg cac Leu Thr Cys Phe His Phe Gly Gly Tyr His His Glu His His Leu His 210 215 220	672
ccg acg gtg cct tgg tgg cgc ctg ccc agc acc cgc acc aag ggg gac Pro Thr Val Pro Trp Trp Arg Leu Pro Ser Thr Arg Thr Lys Gly Asp 225 230 235 240	720
acc gca tga Thr Ala	729

<210> 181
<211> 242
<212> PRT
<213> Paracoccus carotinifaciens E-396

<400> 181

Met Ser Ala His Ala Leu Pro Lys Ala Asp Leu Thr Ala Thr Ser Leu
1 5 10 15

Ile Val Ser Gly Gly Ile Ala Ala Trp Leu Ala Leu His Val His
20 25 30

Ala Leu Trp Phe Leu Asp Ala Ala His Pro Ile Leu Ala Val Ala
35 40 45

Asn Phe Leu Gly Leu Thr Trp Leu Ser Val Gly Leu Phe Ile Ile Ala
50 55 60

His Asp Ala Met His Gly Ser Val Val Pro Gly Arg Pro Arg Ala Asn
65 70 75 80

Ala Ala Met Gly Gln Leu Val Leu Trp Leu Tyr Ala Gly Phe Ser Trp
85 90 95

Arg Lys Met Ile Val Lys His Met Ala His His Arg His Ala Gly Thr
100 105 110

Asp Asp Asp Pro Asp Phe Asp His Gly Gly Pro Val Arg Trp Tyr Ala
115 120 125

Arg Phe Ile Gly Thr Tyr Phe Gly Trp Arg Glu Gly Leu Leu Leu Pro
130 135 140

Val Ile Val Thr Val Tyr Ala Leu Met Leu Gly Asp Arg Trp Met Tyr
145 150 155 160

Val Val Phe Trp Pro Leu Pro Ser Ile Leu Ala Ser Ile Gln Leu Phe
165 170 175

Val Phe Gly Ile Trp Leu Pro His Arg Pro Gly His Asp Ala Phe Pro
180 185 190

Asp Arg His Asn Ala Arg Ser Ser Arg Ile Ser Asp Pro Val Ser Leu
195 200 205

Leu Thr Cys Phe His Phe Gly Gly Tyr His His Glu His His Leu His
210 215 220

Pro Thr Val Pro Trp Trp Arg Leu Pro Ser Thr Arg Thr Lys Gly Asp
225 230 235 240

Thr Ala

<210> 182

<211> 510

<212> DNA

<213> Paracoccus sp. R1534

<220>

<221> CDS

<222> (1)..(507)

<223> Beta-Carotene hydroxylase

<400> 182			
atg agc act tgg gcc gca atc ctg acc gtc atc ctg acc gtc gcc gcg		48	
Met Ser Thr Trp Ala Ala Ile Leu Thr Val Ile Leu Thr Val Ala Ala			
1	5	10	15
atg gag ctg acg gcc tac tcc gtc cat cgg tgg atc atg cat ggc ccc		96	
Met Glu Leu Thr Ala Tyr Ser Val His Arg Trp Ile Met His Gly Pro			
20	25	30	
ctg ggc tgg ggc tgg cat aaa tcg cac cac gac gag gat cac gac cac		144	
Leu Gly Trp Gly Trp His Lys Ser His His Asp Glu Asp His Asp His			
35	40	45	
gcg ctc gag aag aac gac ctc tat ggc gtc atc ttc gcg gta atc tgg		192	
Ala Leu Glu Lys Asn Asp Leu Tyr Gly Val Ile Phe Ala Val Ile Ser			
50	55	60	
atc gtg ctg ttc gcg atc ggc gcg atg ggg tcg gat ctg gcc tgg tgg		240	
Ile Val Leu Phe Ala Ile Gly Ala Met Gly Ser Asp Leu Ala Trp Trp			
65	70	75	80
ctg gcg gtg ggg gtc acc tgc tac ggg ctg atc tac tat ttc ctg cat		288	
Leu Ala Val Gly Val Thr Cys Tyr Gly Leu Ile Tyr Tyr Phe Leu His			
85	90	95	
gac ggc ttg gtg cat ggg cgc tgg ccg ttc cgc tat gtc ccc aag cgc		336	
Asp Gly Leu Val His Gly Arg Trp Pro Phe Arg Tyr Val Pro Lys Arg			
100	105	110	
ggc tat ctt cgt cgc gtc tac cag gca cac agg atg cat cac gcg gtc		384	
Gly Tyr Leu Arg Arg Val Tyr Gln Ala His Arg Met His His Ala Val			
115	120	125	

cat ggc cgc gag aac tgc gtc agc ttc ggt ttc atc tgg gcg ccc tcg	432	
His Gly Arg Glu Asn Cys Val Ser Phe Gly Phe Ile Trp Ala Pro Ser		
130	135	140

gtc gac agc ctc aag gca gag ctg aaa cgc tcg ggc gcg ctg ctg aag 480
 Val Asp Ser Leu Lys Ala Glu Leu Lys Arg Ser Gly Ala Leu Leu Lys
 145 150 155 160

gac cgc gaa ggg gcg gat cgc aat aca tga
Asp Arg Glu Gly Ala Asp Arg Asn Thr
165

<210> 183

<211> 169

<212> PRT

<213> Paracoccus sp. R1534

<400> 183

```

Met Ser Thr Trp Ala Ala Ile Leu Thr Val Ile Leu Thr Val Ala Ala
      5           10          15
1

```

Met Glu Leu Thr Ala Tyr Ser Val His Arg Trp Ile Met His Gly Pro
 20 25 30

Leu Gly Trp Gly Trp His Lys Ser His His Asp Glu Asp His Asp His
 35 40 45

Ala Leu Glu Lys Asn Asp Leu Tyr Gly Val Ile Phe Ala Val Ile Ser
50 55 60

Ile Val Leu Phe Ala Ile Gly Ala Met Gly Ser Asp Leu Ala Trp Trp
55 70 75 80

Leu Ala Val Gly Val Thr Cys Tyr Gly Leu Ile Tyr Tyr Phe Leu His
85 90 95

Asp Gly Leu Val His Gly Arg Trp Pro Phe Arg Tyr Val Pro Lys Arg
100 105 110

Gly Tyr Leu Arg Arg Val Tyr Gln Ala His Arg Met His His Ala Val
115 120 125

His Gly Arg Glu Asn Cys Val Ser Phe Gly Phe Ile Trp Ala Pro Ser
 130 135 140

Val Asp Ser Leu Lys Ala Glu Leu Lys Arg Ser Gly Ala Leu Leu Lys
 145 150 155 160

Asp Arg Glu Gly Ala Asp Arg Asn Thr
 165

<210> 184

<211> 888

<212> DNA

<213> Paracoccus sp. R1534

<220>

<221> CDS

<222> (1)...(885)

<223> farnesyltransferase or geranylgeranyl diphosphate synthase

<400> 184

atg acg ccc aag cag caa ttc ccc cta cgc gat ctg gtc gag atc agg	48
Met Thr Pro Lys Gln Gln Phe Pro Leu Arg Asp Leu Val Glu Ile Arg	
1 5 10 15	

ctg gcg cag atc tcg ggc cag ttc ggc gtg gtc tcg gcc ccg ctc ggc	96
Leu Ala Gln Ile Ser Gly Gln Phe Gly Val Val Ser Ala Pro Leu Gly	
20 25 30	

gcg gcc atg agc gat gcc gcc ctg tcc ccc ggc aaa cgc ttt cgc gcc	144
Ala Ala Met Ser Asp Ala Ala Leu Ser Pro Gly Lys Arg Phe Arg Ala	
35 40 45	

gtg ctg atg ctg atg gtc gcc gaa agc tcg ggc ggg gtc tgc gat gcg	192
Val Leu Met Leu Met Val Ala Glu Ser Ser Gly Gly Val Cys Asp Ala	
50 55 60	

atg gtc gat gcc gcc tgc gcg gtc gag atg gtc cat gcc gca tcg ctg	240
Met Val Asp Ala Ala Cys Ala Val Glu Met Val His Ala Ala Ser Leu	
65 70 75 80	

atc ttc gac gac atg ccc tgc atg gac gat gcc agg acc cgt cgc ggt	288
Ile Phe Asp Asp Met Pro Cys Met Asp Asp Ala Arg Thr Arg Arg Gly	
85 90 95	

cag ccc gcc acc cat gtc gcc cat ggc gag ggg cgc gcg gtg ctt gcg Gln Pro Ala Thr His Val Ala His Gly Glu Gly Arg Ala Val Leu Ala 100 105 110	336
ggc atc gcc ctg atc acc gag gcc atg cgg att ttg ggc gag ggc cgc Gly Ile Ala Leu Ile Thr Glu Ala Met Arg Ile Leu Gly Glu Ala Arg 115 120 125	384
ggc gcg acg ccg gat cag cgc gca agg ctg gtc gca tcc atg tcg cgc Gly Ala Thr Pro Asp Gln Arg Ala Arg Leu Val Ala Ser Met Ser Arg 130 135 140	432
gcf atg gga ccg gtg ggg ctg tgc gca ggg cag gat ctg gac ctg cac Ala Met Gly Pro Val Gly Leu Cys Ala Gly Gln Asp Leu Asp Leu His 145 150 155 160	480
gcc ccc aag gac gcc gcc ggg atc gaa cgt gaa cag gac ctc aag acc Ala Pro Lys Asp Ala Ala Gly Ile Glu Arg Glu Gln Asp Leu Lys Thr 165 170 175	528
ggc gtg ctg ttc gtc gcg ggc ctc gag atg ctg tcc att att aag ggt Gly Val Leu Phe Val Ala Gly Leu Glu Met Leu Ser Ile Ile Lys Gly 180 185 190	576
ctg gac aag gcc gag acc gag cag ctc atg gcc ttc ggg cgt cag ctt Leu Asp Lys Ala Glu Thr Glu Gln Leu Met Ala Phe Gly Arg Gln Leu 195 200 205	624
ggt cgg gtc ttc cag tcc tat gac gac ctg ctg gac gtg atc ggc gac Gly Arg Val Phe Gln Ser Tyr Asp Asp Leu Leu Asp Val Ile Gly Asp 210 215 220	672
aag gcc agc acc ggc aag gat acg ggg cgc gac acc gcc gcc ccc ggc Lys Ala Ser Thr Gly Lys Asp Thr Gly Arg Asp Thr Ala Ala Pro Gly 225 230 235 240	720
cca aag cgc ggc ctg atg gcg gtc gga cag atg ggc gac gtg gcg cag Pro Lys Arg Gly Leu Met Ala Val Gly Gln Met Gly Asp Val Ala Gln 245 250 255	768
cat tac cgc gcc agc cgc gcg caa ctg gac gag ctg atg cgc acc cgg His Tyr Arg Ala Ser Arg Ala Gln Leu Asp Glu Leu Met Arg Thr Arg 260 265 270	816
ctg ttc cgc ggg ggg cag atc gcg gac ctg ctg gcc cgc gtg ctg ccg Leu Phe Arg Gly Gly Gln Ile Ala Asp Leu Leu Ala Arg Val Leu Pro 275 280 285	864
cat gac atc cgc cgc agc gcc tag His Asp Ile Arg Arg Ser Ala 290 295	888

<210> 185

<211> 295

<212> PRT

<213> Paracoccus sp. R1534

<400> 185

Met Thr Pro Lys Gln Gln Phe Pro Leu Arg Asp Leu Val Glu Ile Arg
1 5 10 15

Leu Ala Gln Ile Ser Gly Gln Phe Gly Val Val Ser Ala Pro Leu Gly
20 25 30

Ala Ala Met Ser Asp Ala Ala Leu Ser Pro Gly Lys Arg Phe Arg Ala
35 40 45

Val Leu Met Leu Met Val Ala Glu Ser Ser Gly Gly Val Cys Asp Ala
50 55 60

Met Val Asp Ala Ala Cys Ala Val Glu Met Val His Ala Ala Ser Leu
65 70 75 80

Ile Phe Asp Asp Met Pro Cys Met Asp Asp Ala Arg Thr Arg Arg Gly
85 90 95

Gln Pro Ala Thr His Val Ala His Gly Glu Gly Arg Ala Val Leu Ala
100 105 110

Gly Ile Ala Leu Ile Thr Glu Ala Met Arg Ile Leu Gly Glu Ala Arg
115 120 125

Gly Ala Thr Pro Asp Gln Arg Ala Arg Leu Val Ala Ser Met Ser Arg
130 135 140

Ala Met Gly Pro Val Gly Leu Cys Ala Gly Gln Asp Leu Asp Leu His
145 150 155 160

Ala Pro Lys Asp Ala Ala Gly Ile Glu Arg Glu Gln Asp Leu Lys Thr
165 170 175

Gly Val Leu Phe Val Ala Gly Leu Glu Met Leu Ser Ile Ile Lys Gly
180 185 190

Leu Asp Lys Ala Glu Thr Glu Gln Leu Met Ala Phe Gly Arg Gln Leu

195

200

205

Gly Arg Val Phe Gln Ser Tyr Asp Asp Leu Leu Asp Val Ile Gly Asp
210 215 220

Lys Ala Ser Thr Gly Lys Asp Thr Gly Arg Asp Thr Ala Ala Pro Gly
225 230 235 240

Pro Lys Arg Gly Leu Met Ala Val Gly Gln Met Gly Asp Val Ala Gln
245 250 255

His Tyr Arg Ala Ser Arg Ala Gln Leu Asp Glu Leu Met Arg Thr Arg
260 265 270

Leu Phe Arg Gly Gly Gln Ile Ala Asp Leu Leu Ala Arg Val Leu Pro
275 280 285

His Asp Ile Arg Arg Ser Ala
290 295

<210> 186

<211> 30

<212> DNA

<213> synthetic construct

<400> 186

aaggcctcat atgagcgcac atgccctgcc

30

<210> 187

<211> 28

<212> DNA

<213> synthetic construct

<400> 187

cgggatcctc atgcgggtgtc ccccttgg

28

<210> 188

<211> 30

<212> DNA

<213> synthetic construct

<400> 188

aaggcctcat atgagcaactt gggccgcaat

30

<210> 189

<211> 30

<212> DNA

<213> synthetic construct

<400> 189

aggatcctca tgtattgcga tccgcccctt

30

<210> 190

<211> 52

<212> DNA

<213> synthetic construct

<400> 190

gtgcagcctc aggtcgacat atgcggccgc atccggatcc ctccctctcc ag

52

<210> 191

<211> 52

<212> DNA

<213> synthetic construct

<400> 191

cacgtcggag tccagctgta tacgccggcg taggcctagg gaggaggagg tc

52

<210> 192

<211> 52

<212> DNA

<213> synthetic construct

<400> 192
gtgcaggagg aggtcgacat atgcggccgc atccggatcc ctgaggctcc ag 52

<210> 193

<211> 52

<212> DNA

<213> synthetic construct

<400> 193
cacgtcctcc tccagctgta tacgccggcg taggcctagg gactccgagg tc 52

<210> 194

<211> 52

<212> DNA

<213> synthetic construct

<400> 194
ctggaggcctc aggtcgacat atgcggccgc atccggatcc ctcccttc ac 52

<210> 195

<211> 52

<212> DNA

<213> synthetic construct

<400> 195
gacctcggag tccagctgta tacgccggcg taggcctagg gaggaggacg tg 52

<210> 196

<211> 52

<212> DNA

<213> synthetic construct

<400> 196

ctggaggagg aggtcgacat atgcggccgc atccggatcc ctgaggctgc ac

52

<210> 197

<211> 52

<212> DNA

<213> synthetic construct

<400> 197

gacctcctcc tccagctgta tacgccggcg taggcctagg gactccgacg tg

52.

ATCC

10801 University Blvd • Manassas, VA 20110-2209 • Telephone: 703-365-2700 • FAX: 703-365-2745

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF
THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL FORM

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT ISSUED PURSUANT TO RULE 7.3
AND VIABILITY STATEMENT ISSUED PURSUANT TO RULE 10.2

To: (Name and Address of Depositor or Attorney)

Roche Vitamins Inc.
Attn: Markus Huembelin
340 Kingsland Street
Nutley, NJ 007110-1199

Deposited on Behalf of: Roche Vitamins Inc.

Identification Reference by Depositor:

Paracoccus sp.: R-1506

Patent Deposit Designation

PTA-3431

The deposit was accompanied by: a scientific description a proposed taxonomic description indicated above.

The deposit was received June 5, 2001 by this International Depository Authority and has been accepted.

AT YOUR REQUEST: We will inform you of requests for the strain for 30 years.

The strain will be made available if a patent office signatory to the Budapest Treaty certifies one's right to receive, or if a U.S. Patent is issued citing the strain, and ATCC is instructed by the United States Patent & Trademark Office or the depositor to release said strain.

If the culture should die or be destroyed during the effective term of the deposit, it shall be your responsibility to replace it with living culture of the same.

The strain will be maintained for a period of at least 30 years from date of deposit, or five years after the most recent request for a sample, whichever is longer. The United States and many other countries are signatory to the Budapest Treaty.

The viability of the culture cited above was tested June 27, 2001. On that date, the culture was viable.

International Depository Authority: American Type Culture Collection, Manassas, VA 20110-2209 USA.

Signature of person having authority to represent ATCC:


Tanya Nunnally, Patent Specialist, Patent Depository

Date: June 28, 2001

cc: Kevin C. Hooper
(Ref: Docket or Case No.: C38435/121966)

ATCC

10801 University Blvd • Manassas, VA 20110-2209 • Telephone: 703-365-2700 • FAX: 703-365-2745

**BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF
THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSES OF PATENT PROCEDURE**

INTERNATIONAL FORM

**RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT ISSUED PURSUANT TO RULE 7.3
AND VIABILITY STATEMENT ISSUED PURSUANT TO RULE 10.2**

To: (Name and Address of Depositor or Attorney)

Roche Vitamins Inc.
Attn: Marcus Hucembelin
340 Kingsland Street
Nutley, NJ 07110-1199

Deposited on Behalf of: Roche Vitamins Inc.

Identification Reference by Depositor:
Paracoccus sp.: R114
Paracoccus sp.: R1534

Patent Deposit Designation
PTA-3335
PTA-3336

The deposits were accompanied by: a scientific description, a proposed taxonomic description indicated above. The deposits were received April 24, 2001 by this International Depository Authority and have been accepted.

AT YOUR REQUEST: We will inform you of requests for the strains for 30 years.

The strains will be made available if a patent office signatory to the Budapest Treaty certifies one's right to receive, or if a U.S. Patent is issued citing the strains, and ATCC is instructed by the United States Patent & Trademark Office or the depositor to release said strains.

If the cultures should die or be destroyed during the effective term of the deposit, it shall be your responsibility to replace them with living cultures of the same.

The strains will be maintained for a period of at least 30 years from date of deposit, or five years after the most recent request for a sample, whichever is longer. The United States and many other countries are signatory to the Budapest Treaty.

The viability of the cultures cited above was tested May 7, 2001. On that date, the cultures were viable.

International Depository Authority: American Type Culture Collection, Manassas, VA 20110-2209 USA.

Signature of person having authority to represent ATCC:

Tanya Neunally, Patent Specialist, Patent Depository

Date: May 23, 2001

cc: Kevin C. Hooper
(Ref: Docket or Case No.: CJ8435/121966)

THIS PAGE BLANK (USPTO)

()

()

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 December 2002 (12.12.2002)

PCT

(10) International Publication Number
WO 02/099095 A3

(51) International Patent Classification⁷: C12N 9/02, 9/12, 9/88, 9/90, 15/53, 15/54, 15/60, 15/61, 1/20, 1/21, C12P 7/02, 17/06, C12N 9/10, C12P 23/00

(74) Agents: KELLER, Günter et al.; Lederer & Keller, Prinzregentenstrasse 16, 80538 München (DE).

(21) International Application Number: PCT/EP02/06171

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(22) International Filing Date: 5 June 2002 (05.06.2002)

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(25) Filing Language: English

Published:

(26) Publication Language: English

- with international search report
- with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description

(30) Priority Data:
60/296,299 6 June 2001 (06.06.2001) US

(88) Date of publication of the international search report:
18 December 2003

(71) Applicant (for all designated States except US): ROCHE VITAMINS AG [CH/CH]; 124 Grenzacherstrasse, CH-4070 Basle (CH).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventors; and

(75) Inventors/Applicants (for US only): BERRY, Alan [US/CH]; 99 Mattweg, CH-4144 Arlesheim (CH). BRETZEL, Werner [DE/DE]; 93B Tuellinger Strasse, 79539 Loerrach (DE). HUEMBELIN, Markus [CH/CH]; 2 Wollbacherstrasse, CH-4058 Basle (CH). LOPEZ-ULIBARRI, Rual [MX/CH]; 3 Wasserweg, CH-4334 Sisseln (CH). MAYER, Anne Françoise [LU/CH]; 100 Rennweg, CH-4052 Basle (CH). YELISEEV, Alexei [RU/US]; 1076 Carol Lane, # 118, Lafayette, CA 94549 (US).



WO 02/099095 A3

(54) Title: IMPROVED ISOPRENOID PRODUCTION

(57) Abstract: Isolated polynucleotides encoding polypeptides having the activity of enzymes in the mevalonate pathway, e.g. hydroxymethylglutaryl-CoA reductase, isopentenyl diphosphate isomerase, hydroxymethylglutaryl-CoA synthase, mevolane kinase, phosphomevalonate kinase, or diphosphomevalonate decarboxylase; are provided, useful for recombinantly producing isoprenoid compounds such as carotenoids like phytoene, lycopene, β-carotene, zeaxanthin, canthaxanthin, astaxanthin, adonixanthin, cryptoxanthin, echinenone and adonirubin. Expression vectors, cultured cells, and methods of making isoprenoid compounds are also provided.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 02/06171

A. CLASSIFICATION OF SUBJECT MATTER					
IPC 7	C12N9/02	C12N9/12	C12N9/88	C12N9/90	C12N15/53
	C12N15/54	C12N15/60	C12N15/61	C12N1/20	C12N1/21
	C12P7/02	C12P17/06	C12N9/10	C12P23/00	

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, SEQUENCE SEARCH, BIOSIS, WPI Data, PAJ, FSTA, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	<p>DATABASE EM_PRO 'Online!' EMBL; 2 May 2002 (2002-05-02) HUEMBELIN, M. ET AL.: "Paracoccus zeaxanthinifaciens mvaA, idi, hcs, mvk, pmk, and mvd gene." retrieved from EBI, accession no. AJ431696 Database accession no. PZE431696 XP002232853 abstract</p> <p>---</p> <p style="text-align: center;">-/--</p>	1, 6

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

26 May 2003

Date of mailing of the international search report

07.07.2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl
 Fax: (+31-70) 340-3016

Authorized officer

De Kok, A

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 02/06171

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	<p>DATABASE EM_PRO 'Online! EMBL; 2 May 2002 (2002-05-02) HUEMBELIN, M. ET AL.: "Paracoccus zeaxanthinifaciens xseB and ispA gene" retrieved from EBI, accession no. PZE431697 Database accession no. AJ431697 XP002242421 abstract</p> <p>---</p> <p>DATABASE EM_PRO 'Online! EMBL; 2 May 2002 (2002-05-02) HUEMBELIN, M.: "Paracoccus zeaxanthinifaciens phaA gene, phaB gene, ORF1 and ddsA gene" retrieved from EBI, accession no. PZE431695 Database accession no. AJ431695 XP002242422 abstract</p> <p>---</p> <p>EISENREICH WOLFGANG ET AL: "Biosynthesis of zeaxanthin via mevalonate in Paracoccus species strain PTA-3335. A product-based retrobiosynthetic study." JOURNAL OF ORGANIC CHEMISTRY, vol. 67, no. 3, 8 February 2002 (2002-02-08), pages 871-875, XP002242419 February 8, 2002 ISSN: 0022-3263 the whole document</p> <p>---</p> <p>US 3 891 504 A (SCHOCHER ARNO JOHANNES ET AL) 24 June 1975 (1975-06-24) cited in the application the whole document</p> <p>---</p> <p>WO 00 01650 A (DCV INC) 13 January 2000 (2000-01-13) page 8, line 1 - line 24 page 14, line 12 -page 20, line 27</p> <p>---</p> <p>EP 0 747 483 A (HOFFMANN LA ROCHE) 11 December 1996 (1996-12-11) the whole document, especially page 11, lines 10-26.</p> <p>& US 6 087 152 A 11 July 2000 (2000-07-11) cited in the application</p> <p>---</p>	2,3,7,9 4,5,8,9 16 16 14 1,6, 9-13,15 4,8,16 1,6, 9-13,15

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 02/06171

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YABUTANI TETSUYA ET AL: "Analysis of beta-ketothiolase and acetoacetyl-CoA reductase genes of a methylotrophic bacterium, <i>Paracoccus denitrificans</i> , and their expression in <i>Escherichia coli</i> ." <i>FEMS MICROBIOLOGY LETTERS</i> , vol. 133, no. 1-2, 1995, pages 85-90, XP002242420 ISSN: 0378-1097 the whole document -& DATABASE EM_PRO 'Online' EMBL; 20 February 1995 (1995-02-20) YABUTANI, T. ET AL.: "Paracoccus denitrificans genes for beta-ketothiolase and acetoacetyl-CoA reductase" retrieved from EBI, accession no. PDPHAA Database accession no. D49362 XP002242423 abstract -& DATABASE SWALL 'Online' 1 October 1996 (1996-10-01) YABUTANI, T. ET AL.: "Acetoacetyl-CoA-thiolase (phaA)" retrieved from EBI, accession no. THIL-PARDE Database accession no. P54810 XP002242424 abstract ---	4,5,8-13
X	--- DATABASE SWALL 'Online' 1 October 1996 (1996-10-01) YABUTANI, T. ET AL.: "Acetoacetyl-CoA-reductase (PhaB)" retrieved from EBI, accession no. PHAB_PARDE Database accession no. P50204 XP002242425 abstract ---	4,5,8,9
Y	WILDING E IMOGEN ET AL: "Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci." <i>JOURNAL OF BACTERIOLOGY</i> , vol. 182, no. 15, August 2000 (2000-08), pages 4319-4327, XP002232852 ISSN: 0021-9193 abstract ---	1,6, 9-13,15
Y	EP 1 072 683 A (KYOWA HAKKO KOGYO KK) 31 January 2001 (2001-01-31) page 3, paragraph 20 -page 12, paragraph 160 ---	2,7
A	---	3,9-15
	-/-	

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 02/06171

C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 99 06586 A (YISSUM RES DEV CO) 11 February 1999 (1999-02-11) the whole document ---	1,6, 9-13,15, 16
Y		2,7
T	HUMBELIN M ET AL: "Genetics of isoprenoid biosynthesis in Paracoccus zeaxanthinifaciens" GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES, ELSEVIER SCIENCE PUBLISHERS, BARKING, GB, vol. 297, no. 1-2, 4 September 2002 (2002-09-04), pages 129-139, XP004388023 ISSN: 0378-1119 the whole document ---	1,6,9-16
T	DATABASE SWALL 'Online! 1 October 2002 (2002-10-01) HUEMBELIN, H.: "Fragment of deoxyxylulose 5-phosphate synthase" retrieved from EBI, accession no. Q8L1H7 Database accession no. Q8L1H7 XP002242426 abstract -----	3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP 02/06171

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a). ()

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

As a result of the prior review under R. 40.2(e) PCT,
no additional fees are to be refunded.

1. As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee. ()

3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
- No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1, 6, 14-16 completely and claims 9-13 partially

1.1. Claims: 1, 6, 15 completely and 9-13 partially
An isolated polypeptide selected from a polypeptide comprising the aminoacid sequence shown in SEQ.ID.No. 43, 45, 47, 49, 51 or 53; a polynucleotide encoding said polypeptides comprising SEQ.ID.No. 24 and their use.

1.2. Claim : 14
A method of making a carotenoid-producing cell

1.3. Claim : 16
A microorganism of the genus Paracoccus

2. Claims: 2 completely and 7, 9-13 partially

An isolated polypeptide having the aminoacid sequence shown in SEQ.ID.No. 159; a polynucleotide encoding said polypeptide comprising SEQ.ID.No. 157 and their use.

3. Claims: 3 completely and 7, 9-13 partially

An isolated polypeptide having the aminoacid sequence shown in SEQ.ID.No. 160; a polynucleotide encoding said polypeptide comprising SEQ.ID.No. 157 and their use.

4. Claims: 4, 5, 8 completely and 9-13 partially

An isolated polypeptide having the aminoacid sequence shown in SEQ.ID.No. 178 resp. 179; a polynucleotide encoding said polypeptides comprising SEQ.ID.No. 177 and their use.

Please note that all inventions mentioned under item 1, although not necessarily linked by a common inventive concept, could be searched without effort justifying an additional fee.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 02/06171

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
US 3891504	A	24-06-1975		CH 549094 A AT 304414 B BE 770744 A1 CA 950848 A1 DE 2138000 A1 ES 393771 A1 FR 2099367 A5 GB 1318828 A IT 1051240 B NL 7110401 A SE 377135 B TR 17137 A ZA 7104028 A	15-05-1974 10-01-1973 31-01-1972 09-07-1974 10-02-1972 01-09-1973 10-03-1972 31-05-1973 21-04-1981 02-02-1972 23-06-1975 25-04-1974 29-03-1972
WO 0001650	A	13-01-2000		AU 4863099 A CA 2331343 A1 CN 1330621 T EP 1095002 A1 JP 2002519049 T WO 0001650 A1 US 6410755 B1 US 6531303 B1	24-01-2000 13-01-2000 09-01-2002 02-05-2001 02-07-2002 13-01-2000 25-06-2002 11-03-2003
EP 0747483	A	11-12-1996		EP 0747483 A2 JP 9023888 A US 6124113 A US 6207409 B1 US 2002147371 A1 US 6087152 A	11-12-1996 28-01-1997 26-09-2000 27-03-2001 10-10-2002 11-07-2000
EP 1072683	A	31-01-2001		AU 3169999 A CA 2325798 A1 EP 1072683 A1 CN 1305529 T WO 9953071 A1 JP 2000300256 A JP 2000300257 A	01-11-1999 21-10-1999 31-01-2001 25-07-2001 21-10-1999 31-10-2000 31-10-2000
WO 9906586	A	11-02-1999		US 5935808 A AU 749302 B2 AU 8575198 A EP 1005565 A1 JP 2001512030 T WO 9906586 A1	10-08-1999 20-06-2002 22-02-1999 07-06-2000 21-08-2001 11-02-1999